

Investigating high-affinity non-covalent protein-ligand interaction via variants of streptavidin



Claire Elizabeth Chivers
Department of Biochemistry
Somerville College, University of Oxford

**A thesis submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy, University of Oxford
Trinity Term 2011**

*This thesis is dedicated to my parents
Christine and Cyril Chivers*

Declaration

The work described in this thesis was carried out between April 2008 and February 2011, in the laboratory of Dr. Mark Howarth at the Biochemistry Department of the University of Oxford. All the work described within this thesis is entirely my own unless otherwise stated. This work has not been submitted previously for any other degree at the University of Oxford or any other university. Publications arising from this work are detailed prior to the document contents.

Claire E. Chivers

Trinity 2011

Acknowledgments

I would first like to briefly acknowledge the people who have helped me get to the DPhil stage. My 3 years at Cardiff University, with inspiring lecturers, excellent learning resources and support at the Biosciences department, made a lasting impression on me. My 2 summers at the University of Oulu's Biotechnology Laboratory in Sotkamo confirmed my love of research and was the source of a large proportion of my motivation to continue onto the DPhil.

The last 4 years have been a huge learning curve for me. I am very grateful for being given the opportunity to work in the motivational research atmosphere at Oxford as part of the BBSRC Chemical Biology and Molecular Biochemistry DPhil programme. Scientifically, I have learnt how to conduct research on a limited budget, as well as the value of both perseverance and serendipity in research. The inter-group collaboration encouraged by the Biochemistry department has been particularly rewarding. My thanks go to all members of the Sherratt group for their help and advice on FtsK. I am also particularly grateful that Dr Ed Lowe allowed me to work alongside him whilst learning x-ray crystallography. My supervisor Dr Mark Howarth constantly amazes me with his dedication to research. He has constantly pushed me to strive on, and as a result of his encouragement and direction, I am very happy with what I have learnt and achieved in his lab during my DPhil. I appreciate Mark always being available for discussion / advice on the latest results / papers and his energy and enthusiasm for research is truly inspirational.

I have also learnt the value of being surrounded by supportive friends and colleagues. My first 6 months in Oxford were spent in Prof. Jim McDonnell's lab, and I often wonder whether I would have remained at Oxford had I not started in his friendly lab. During this time, I met Dr Susmita Borthakur – a huge thank you goes to you for your valued friendship over the years! A big thank you also goes to the past and current members of the Howarth group and 40-007 lab who have made the long hours much more endurable. My particular thanks go to Dr Denis Krndija and Dr Paola Bisicchia for their tremendous support and frequent hugs! Outside the lab, my thanks go to Somerville College for their graduate scholarship and to the lodge porters, who welcomed me into their motley crew for over a year and kept me sane and well-fed! I am also extremely grateful to have been working with such a fantastic group of people in the science policy centre at the Royal Society during some rather hectic thesis-writing months. I think they, like many other people I have known over the years, might be unaware of the positive impact they have had on me – a huge thank you goes to them, and to the others.

Last but by no means least, I have learnt that there is just no end to my family's love and support. Auntie Pam and Uncle Trevor, Auntie Mary and Uncle Alun, Becky (and Sam!) all deserve a big thank you for being so supportive. My parents' truly unconditional love has enabled me to be brave and try new things without fear of failing, and I would be nothing without them. This DPhil, together with anything else I have achieved in my life, is purely a result of their unceasing support, encouragement, energy and love. Thank you, Mam and Dad, this is for you.

Abstract

Investigating high-affinity non-covalent protein-ligand interaction via variants of streptavidin

Claire Elizabeth Chivers – Somerville College – University of Oxford

Submitted for the degree of Doctor of Philosophy in Trinity Term 2011

The *Streptomyces avidinii* protein streptavidin binds the small molecule biotin (vitamin H / B₇) with extraordinary stability, resulting in the streptavidin-biotin interaction being one of the strongest non-covalent interactions known in nature ($K_d \sim 10^{-14}$ M). The stable and rapid biotin-binding, together with high resistance to heat, pH and proteolysis, has given streptavidin huge utility, both *in vivo* and *in vitro*. Accordingly, streptavidin has become a widely used tool in many different biotechnological applications. Streptavidin has also been the subject of extensive research efforts to glean insights into this paradigm for a high-affinity interaction, with over 200 mutants of the protein reported to date.

Despite the high stability of the streptavidin-biotin interaction, it can and does fail under certain experimental conditions. For example, streptavidin-biotin dissociation is accelerated by an increased temperature or lower pH (conditions often encountered in cellular imaging experiments), and by mechanical stress, such as the shear force arising from fluid flow (encountered when streptavidin is used as a molecular anchor in biosensor chips and arrays).

This study details efforts made at increasing further the utility of streptavidin, by increasing the stability of biotin and biotin-conjugate binding. A rational site-directed mutagenesis approach was used to create 27 mutants, with eight of these mutants possessing higher-stability biotin-binding. The most stable biotin-binding mutant was named traptavidin and was extensively characterised. Kinetic characterisation revealed traptavidin had a decreased dissociation rate from biotin and biotin-conjugates when compared to wildtype streptavidin, at both neutral pH and pH 5. Atomic force microscopy and molecular motor displacement assays revealed the traptavidin-biotin interaction possessed higher mechanical stability than the streptavidin-biotin interaction. Cellular imaging experiments revealed the non-specific cell binding properties of streptavidin were unchanged in traptavidin. X-ray crystallography was also used to generate structures of both apo- and biotin-bound traptavidin at 1.5 Å resolution. The structures were analysed in detail and compared to the published structures of streptavidin, revealing the characteristics of traptavidin arose from the mutations stabilising a flexible loop over the biotin-binding pocket, as well as reducing the conformational change on biotin-binding to traptavidin.

Traptavidin has the potential to replace streptavidin in many of its diverse applications, as well as providing an insight into the nature of ultra-stable non-covalent interactions.

Publications

Research Papers

Chivers, C.E., Koner, A.L., Lowe, E.D. and Howarth, M. 2011. How the biotin-streptavidin interaction was made even stronger: investigation via crystallography and a chimaeric tetramer. *Biochemical Journal* **435** pp 55-63

Crozat, E., Meglio, A., Allemand, J.F., Chivers, C.E., Howarth, M., Venien-Bryan, C., Grainge, I. and Sherratt, D.J. 2010. Separating speed and ability to displace roadblocks during DNA translocation by FtsK. *EMBO J.* **29** pp 1423-1433

Chivers, C.E., Crozat, E., Chu, C., Moy, V.T., Sherratt, D.J. and Howarth, M. 2010. A streptavidin variant with slower biotin dissociation and increased mechanostability. *Nature Methods* **7** pp 391-393

Intellectual Property

Chivers, C.E. and Howarth, M. 2009
Streptavidin mutant proteins. *International Application PCT/GB2010/002006*

Abbreviations

A ₂₈₀	Absorbance at 280 nm
AAA+	<u>A</u> TPases <u>A</u> ssociated with various cellular <u>A</u> ctivities
AFM	Atomic force microscopy
Amp	Ampicillin
AMPA	α -amino-3-hydroxy-5-methyl-4-isoxazolepropionate
AP	Acceptor peptide (GLNDIFEAQKIEWHE)
ATP	Adenosine triphosphate
Av	Avidin
AVR	Avidin-related protein
B4F	Biotin-4-fluorescein
BFP	Biomembrane force probe
bp	Base pair
BirA	Biotin ligase
BNP	Biotinyl <i>p</i> -nitrophenyl ester
BSA	Bovine serum albumin
COS7	African Green Monkey SV40-transfected kidney fibroblast cell line
D4	Tetrameric 'dead' SA, incapable of binding biotin.
ddH ₂ O	Double distilled water
DHFR	Dihydrofolate reductase
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
DPI	Dispersion precision indicator
DSC	Differential scanning calorimetry
dsDNA	Double stranded deoxyribonucleic acid
DTT	Dithiothreitol
<i>E. coli</i>	<i>Escherichia coli</i>
EDC	1-Ethyl-3-[3-dimethylaminopropyl]carbodiimide hydrochloride
EDTA	Ethylenediamine tetraacetic acid
ELISA	Enzyme-linked immunosorbent assay
ER	Endoplasmic reticulum
ExPASy	<u>E</u> xpert <u>P</u> rotein <u>A</u> nalysis <u>S</u> ystem
FIAsH	Fluorescein arsenical helix binder
FP	Fluorescent protein
GFP	Green fluorescent protein
GuHCl	Guanidine hydrochloride
HABA	2-(4'-hydroxyazobenzene)benzoic acid
H/D	Hydrogen/deuterium
HRP	Horseradish peroxidase
IB	2-iminobiotin
IF2	Initiation factor 2
IF2-SA	His-tagged streptavidin with domain 1 of IF2 at its N terminus
IF2-TR	His-tagged traptavidin with domain 1 of IF2 at its N terminus
IGF1R	Insulin-like growth factor-1 receptor
IMAC	Immobilised metal affinity chromatography
IPTG	Isopropyl- β -D-thiogalactopyranoside

ITC	Isothermal titration calorimetry
kb	kilobases
K_d	Equilibrium dissociation constant; k_{off}/k_{on}
kDa	kilo Daltons
k_{off}	Dissociation rate
k_{on}	Association rate
KOPS	FtsK orienting / polarising sequence (GGGNAGGG)
L3/4	Loop connecting β -strands 3 and 4 (residues 45 to 52 in SA/Tr)
LB	Luria Broth
MBP	Maltose-binding protein
MR	Molecular replacement
MWCO	Molecular weight cut-off
n/a	Not applicable
NCL	Native chemical ligation
ND	Not determined
Ni-NTA	Nickel-nitrilotriacetic acid
OD ₆₀₀	Optical density at 600 nm
O/N	Overnight
<i>P. aeruginosa</i>	<i>Pseudomonas aeruginosa</i>
PBS	Phosphate buffered saline
PBS/Mg	Phosphate buffered saline with 5 mM MgCl ₂
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PEG	Poly(ethylene glycol)
pI	Isoelectric point
PMSF	Phenylmethylsulfonyl fluoride
QD	Quantum dot
RNA	Ribonucleic acid
rmsd	Root-mean-square deviation
RT	Room temperature
SA	Streptavidin
S.D.	Standard deviation
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate – polyacrylamide gel electrophoresis
TAE	Tris-Acetate-EDTA
TBE	Tris-Boric acid-EDTA
T_m	Transition temperature
Tr	Traptavidin (S52G R53D mutant of streptavidin)
Tr1D3	Monovalent traptavidin
TRT	Targeted Radionuclide Therapy
WT	Wildtype

List of Contents

Declaration	i
Acknowledgements	ii
Abstract	iii
Wordle	iv
Publications	v
Abbreviations	vi
List of Contents	viii
Detailed List of Contents	ix
Chapter 1: Introduction	
1.1 Biotin-binding proteins	1
1.2 The avidin superfamily	12
1.3 Streptavidin engineering	30
1.4 Applications of streptavidin	44
1.5 Alternative tagging strategies for bioorthogonal chemistry	53
1.6 Aims of this thesis	63
Chapter 2: General Materials and Methods	65
Chapter 3: Production and Characterisation of Higher-Stability Streptavidin Mutants	
3.1 Introduction	70
3.2 Materials and methods	75
3.3 Results	81
3.4 Discussion	94
Chapter 4: Characterisation of Traptavidin	
4.1 Introduction	100
4.2 Materials and methods	114
4.3 Results	130
4.4 Discussion	149
Chapter 5: The Structures of Apo-Traptavidin and Biotin-Bound Traptavidin	
5.1 Introduction	158
5.2 Materials and methods	162
5.3 Results	171
5.4 Discussion	200
Chapter 6: Conclusions	
6.1 A summary of the results of this work	206
6.2 Streptavidin's future prospects	209
Appendices	
A1: Biotin-4-fluorescein binding data for all streptavidin mutants made in this study	213
A2: Optimising streptavidin / traptavidin expression and purification protocols	217
References	250
Publications	275

Detailed List of Contents

Declaration	i
Acknowledgements	ii
Abstract	iii
Wordle	iv
Publications	v
Abbreviations	vi
List of Content	viii
Detailed List of Contents	ix
Chapter 1: Introduction	
1.1 Biotin-binding proteins	1
1.1.1 Discovery of streptavidin & avidin	1
1.1.2 Role of biotin in biological systems	2
1.1.3 Theories on biological role of streptavidin / avidin	5
1.1.4 Biotin analogues / derivatives	6
1.1.5 Other streptavidin binders	10
1.2 The avidin superfamily	12
1.2.1 Streptavidin structure	12
1.2.2 The biotin binding site & the basis for high-affinity binding	14
1.2.3 Other biotin-binders in the avidin superfamily	21
1.2.3.1 Avidin	23
1.2.3.2 Avidin-related proteins	26
1.2.3.3 Avidin superfamily members isolated from bacteria	26
1.2.3.4 Avidin superfamily members isolated from other species	28
1.3 Streptavidin engineering	30
1.3.1 Rational site-directed mutagenesis	32
1.3.1.1 Rational site-directed mutagenesis of SA	34
1.3.2 Combining rational mutagenesis with directed evolution	39
1.3.2.1 A tandem rational mutagenesis / directed evolution approach to SA engineering	42
1.4 Applications of streptavidin	44
1.4.1 Targeted radionuclide therapy	45
1.4.2 Cellular imaging	47
1.4.2.1 The BirA / AP system	48
1.4.3. Detection and diagnostics	50
1.4.4. Nanofabrication	52
1.5 Alternative tagging strategies for bioorthogonal chemistry	53
1.5.1 Click chemistry	54
1.5.2 Native chemical ligation	55
1.5.3 Antibody-antigen binding	57
1.5.4 HaloTag	59
1.5.5 Dihydrofolate reductase	60
1.5.6 FIAsh	61
1.6 Aims of this thesis	63

Chapter 2: General Materials and Methods

2.1 Sources of reagents and enzymes	65
2.2 Culture media	65
2.3 Competent bacterial strains and transformation	66
2.4 Site-directed mutagenesis	67
2.5 Agarose gel electrophoresis	67
2.6 Protein expression and purification	68
2.7 Electrophoretic separation of proteins	68
2.8 Figure presentation	69

Chapter 3: Production and Characterisation of Higher-Stability Streptavidin Mutants

3.1 Introduction	70
3.1.1 The need for a higher-stability biotin-binder	71
3.1.2 Aims of this chapter	73
3.2 Materials and methods	75
3.2.1 Site-directed mutagenesis	75
3.2.2 Protein expression and purification	76
3.2.3 Biotin-4-fluorescein off-rate assay	78
3.2.4 Biotin-4-fluorescein on-rate assay	79
3.2.5 Thermostability assay	80
3.3 Results	81
3.3.1 Protein production	81
3.3.2 Protein expression and purification	82
3.3.3 Biotin-4-fluorescein off-rates	85
3.3.4 Biotin-4-fluorescein on-rates	88
3.3.5 Thermostability	90
3.3.6 Summary	93
3.4 Discussion	94

Chapter 4: Characterisation of Traptavidin

4.1 Introduction	100
4.1.1 Visualising cell surface IGF1R with the BirA / AP system	101
4.1.2 FtsK displacement assays	102
4.1.2.1 FtsK discovery	102
4.1.2.2 FtsK cellular role	102
4.1.2.3 FtsK structure	105
4.1.2.4 Constructs used in this study	108
4.1.2.5 Investigating FtsK	110
4.1.3 Atomic force microscopy	112
4.2 Materials and methods	114
4.2.1 Protein production, expression and purification	114
4.2.2 Biotin-4-fluorescein off-rate assays	114
4.2.2.1 Comparing SA, Tr and avidin B4F off-rates	114
4.2.2.2 Troubleshooting the effect of DMSO on fluorescence	115
4.2.2.3 Long-term B4F off-rate assay (pH 7.4)	115
4.2.2.4 B4F off-rate assay (pH 5)	116
4.2.3 [³ H]biotin assays	117
4.2.3.1 [³ H]biotin off-rate assay	117
4.2.3.2 [³ H]biotin on-rate assay	118
4.2.3.3 Equilibrium dissociation constant calculation	118
4.2.4 Thermostability assays	119
4.2.5 Cell imaging with the BirA / AP system	120
4.2.5.1 Cell culture	120
4.2.5.2 Alexa Fluor® 555 labelling of SA and Tr	120
4.2.5.3 Plasmid construction	120
4.2.5.4 Transfection and biotinylation	121
4.2.5.5 Cell imaging	121
4.2.6 FtsK displacement assays	122
4.2.6.1 Displacement with <i>Pseudomonas aeruginosa</i> FtsK	122
4.2.6.1.1 Initial PAK4 displacement assay	123
4.2.6.1.2 Investigating effect of PAK4 translocation time	124
4.2.6.1.3 Investigating effect of PAK4 concentration	124
4.2.6.2 Displacement with <i>Escherichia coli</i> FtsK	125
4.2.7 Atomic force microscopy	127
4.3 Results	130
4.3.1 Biotin-4-fluorescein off-rate assays	130
4.3.2 [³ H]biotin assays	135
4.3.3 Thermostability assays	138
4.3.4 Live cell imaging	141
4.3.5 FtsK displacement assays	143
4.3.6 Atomic force microscopy	146
4.4 Discussion	149

Chapter 5: The Structures of Apo-Traptavidin and Biotin-Bound Traptavidin	
5.1 Introduction	158
5.2 Materials and methods	162
5.2.1 Expression and purification of proteins used in this study	162
5.2.1.1 Tetravalent traptavidin expression and purification	162
5.2.1.2 Monovalent traptavidin expression and purification	163
5.2.1.3 Tetravalent 'dead' streptavidin expression and purification	163
5.2.2 Crystallisation	165
5.2.2.1 Initial screening of crystallisation conditions	165
5.2.2.2 Optimisation of crystallisation conditions	165
5.2.2.3 Final crystallisation conditions used: apo-Tr	166
5.2.2.4 Final crystallisation conditions used: biotin-Tr	166
5.2.2.5 Diffraction data collection	167
5.2.2.6 Structure solution and refinement: apo-Tr	167
5.2.2.7 Structure solution and refinement: biotin-Tr	168
5.2.3 Structure analysis	170
5.2.4 Biotin-4-fluorescein off-rate assay	170
5.2.5 Thermostability analysis of proteins	170
5.3 Results	171
5.3.1 Traptavidin protein purification	171
5.3.2 Crystallisation	172
5.3.2.1 Initial screening & optimisation of crystallisation conditions	172
5.3.2.2 Final crystallisation conditions used for data collection	174
5.3.2.3 Solution of apo-Tr and biotin-Tr structures	175
5.3.3 Structural analysis	178
5.3.3.1 Tertiary and quaternary structure of traptavidin	178
5.3.3.2 L3/4 conformation in traptavidin	181
5.3.3.3 Conformational change on biotin binding	183
5.3.3.4 Flexibility analysis of streptavidin and traptavidin	185
5.3.3.5 Conformation of bound biotin in streptavidin and traptavidin	189
5.3.3.6 Hydrogen-bonding network in traptavidin	191
5.3.3.7 Intersubunit contacts	196
5.4 Discussion	200
Chapter 6: Conclusions	206
6.1 A summary of the results of this work	206
6.2 Streptavidin's future prospects	209

Appendices

A1: Biotin-4-fluorescein binding data for all streptavidin mutants made in this study	213
A2: Optimising streptavidin / traptavidin expression and purification protocols	
A2.1 Introduction	217
A2.1.1 Streptavidin expression systems	217
A2.1.1.1 Expression in bacteria	218
A2.1.1.1.1 Soluble expression in bacteria	219
A2.1.2 Aim of this chapter	223
A2.2 Materials and methods	224
A2.2.1 IF2-SA plasmid	224
A2.2.2 Expression trials of IF2-SA	224
A2.2.3 Ni-NTA purification of soluble IF2-SA from BL21 (DE3) RIPL cells	225
A2.2.4 Enzymatic digestion of IF2-SA	225
A2.2.5 Enzymatic digestion of soluble fraction of IF2-SA cell pellet	226
A2.2.6 2-iminobiotin affinity chromatography of core-SA from IF2-SA digest	226
A2.2.7 Site-directed mutagenesis to make IF2-Tr	227
A2.2.8 IF2-Tr expression trials	228
A2.2.9 Optimised IF2-Tr expression and purification	229
A2.2.10 Ni-NTA purification of soluble IF2-Tr from B834 (DE3) pLysS cells	229
A2.2.11 Enzymatic digestion of soluble fraction of IF2-Tr cell pellet	230
A2.3 Results	232
A2.3.1 Expression of IF2-SA	232
A2.3.2 Purification of IF2-SA from BL21 (DE3) RIPL cells	233
A2.3.3 Enzymatic digestion of IF2-SA	235
A2.3.4 Putting Tr mutations into IF2-SA plasmid	239
A2.3.5 IF2-Tr expression trials	239
A2.3.6 IF2-Tr expression and purification	243
A2.3.7 Enzymatic digestion of IF2-Tr	245
A2.4 Discussion	247
References	250
Publications	275

Chapter 1: Introduction

1.1 Biotin-binding proteins

1.1.1 Discovery of streptavidin & avidin

The story of biotin-binding proteins and their use in biotechnology begins in 1927, when rats fed on a diet consisting of only raw egg whites developed dermatitis and hair loss, followed by dramatic weight loss and eventually death (Boas, 1927). These detrimental effects of a diet rich in egg white were also seen in chicks (Ringrose et al., 1931), guinea-pigs, rabbits and monkeys (Lease et al., 1937). This so-called 'egg-white injury' could be reversed by including other foodstuffs in the diet, such as liver, yeast and cow's milk, which were hypothesised to contain a certain 'protective factor', which was named vitamin H or biotin (Gyorgy, 1939; Gyorgy et al., 1940). This led to the discovery in 1940 that 'egg white injury' was caused by a deficiency in biotin (Eakin et al., 1940a; Eakin et al., 1940b), and in 1941 the component in egg white responsible for the sequestering of biotin was isolated and named avidin (Av) (Eakin et al., 1941). This was the first biotin-binding protein identified.

Over 20 years later, a second biotin-binding protein, streptavidin (SA), was identified. In 1963-64, a 60 kDa component of the antibiotic MSD-235, produced by *Streptomyces* bacteria, was found to have biotin-binding ability (Chaiet et al., 1963; Tausig and Wolf, 1964); the protein was isolated, characterised and named streptavidin after its comparable biotin-binding properties to avidin (Chaiet and Wolf, 1964).

1.1.2 Role of biotin in biological systems

Biotin is a 244 Da water-soluble vitamin, also called vitamin H / B₇, that is required by all living organisms for healthy cellular growth and function. Biotin can be synthesised by plants, fungi and bacteria; however, animals obtain biotin from dietary sources (as well as from synthesis by the intestinal bacterial flora). Biotin is an essential cofactor in carboxylase enzymes that play critical roles in metabolism (Moss and Lane, 1971). The number of enzymes dependent on biotin as a cofactor varies between organisms, but in all such enzymes, biotin acts as the carrier of a carboxyl group from a donor substrate to an acceptor substrate, in carboxylation, decarboxylation and transcarboxylation reactions. For example, in mammals there are four biotin-dependent carboxylases: acetyl-CoA carboxylase (fatty acid synthesis and lipogenesis), pyruvate carboxylase (gluconeogenesis), propionyl-CoA carboxylase (isoleucine, methionine, threonine and valine catabolism) and 3-methylcrotonyl-CoA carboxylase (leucine catabolism) (Samols et al., 1988).

Biotin has a bicyclic structure (an ureido ring fused to a tetrahydrothiophene ring) with a valeric acid side chain (Figure 1) (du Vigneaud, 1942).

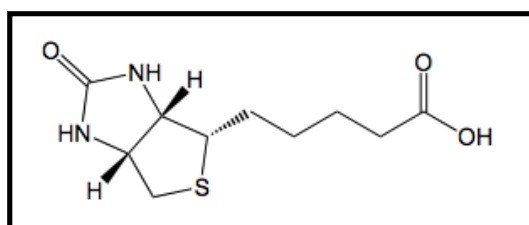


Figure 1. Structure of biotin. The 244 Da vitamin consists of an ureido ring fused to a tetrahydrothiophene ring, with a valeric acid side chain.

The covalent linkage of biotin to the carboxylases is catalysed by a biotin protein ligase enzyme, which is called holocarboxylase synthetase in humans (Leon-Del-Rio et al., 1995) and BirA in bacteria (Barker and Campbell, 1981). These enzymes catalyse the covalent attachment of biotin via its valeryl carboxyl group to the ϵ -amino group of a lysine residue in the almost universally conserved biotin acceptor sequence (A/V)MKM (Samols et al., 1988), producing a biocytin residue (Figure 2). The amide linkage between biotin and the lysine residue in biocytin can be cleaved by biotinidase, liberating free biotin from the carboxylase enzymes (as well as from biotinylated dietary proteins) (Thoma and Peterson, 1954; Wolf et al., 1985). The free biotin can then be reutilized; this is the so-called biotin cycle (Hymes and Wolf, 1996).

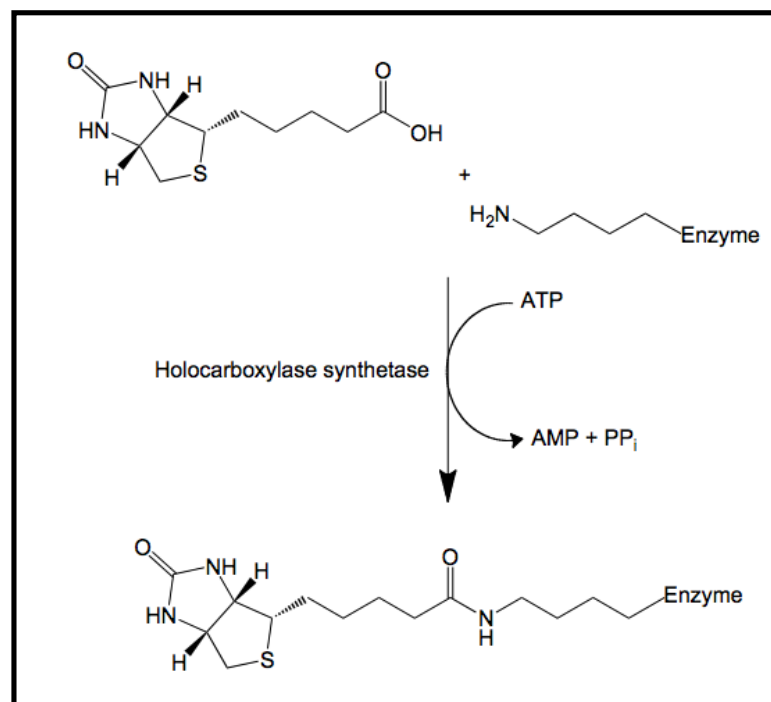


Figure 2. The covalent attachment of biotin to carboxylases is catalysed by holocarboxylase synthetase in mammals and BirA in bacteria. Biotin is attached to a lysine residue in the conserved sequence (A/V)MKM in the carboxylase, producing a biocytin residue.

Once attached to the carboxylase, biotin carries the carboxyl group that is to be transferred via a covalent attachment to its 1'-N ureido nitrogen atom, which requires ATP and converts biotin to 1'-N-carboxy-d-biotin. In the second step of the reaction, the carboxyl group is transferred to the acceptor molecule, producing a free biotin-containing carboxylase that is ready to accept another carboxyl group (Figure 3).

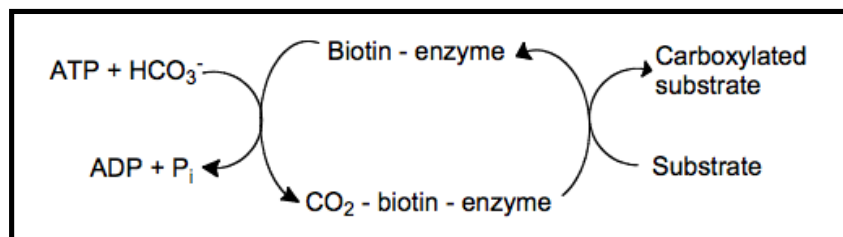


Figure 3. The biotin cycle. The biotin cofactor in carboxylase enzymes carries a carboxyl group in the form of 1'-N-carboxy-biotin. This carboxyl group can then be transferred to a substrate molecule, regenerating a carboxylase enzyme capable of accepting another carboxyl group.

As well as its catalytic role as a carboxylase cofactor, biotin may have a role in regulating gene function (Zempleni, 2005), possibly through covalent attachment to histones, although this has been questioned (Healy et al., 2009).

In humans, biotin deficiency can be a result of diet (a low consumption of biotin-containing foods or excessive consumption of raw egg whites), gastrointestinal malabsorption of biotin due to long-term antibiotic or anticonvulsant use, a deficiency in the holocarboxylase synthetase enzyme that attaches biotin to its dependent enzymes, or a deficiency in biotinidase that releases bound biotin from dietary proteins (Jen and Yan, 2010). The

symptoms of biotin deficiency include dermatitis, hair loss, lethargy, seizures, hearing loss and vision impairment (Jen and Yan, 2010).

1.1.3 Theories on biological role of streptavidin / avidin

Both streptavidin and avidin can bind biotin with an extremely high affinity (avidin's K_d is reported to be 10^{-16} M whilst streptavidin's K_d is reported to be in the range 10^{-14} - 10^{-15} M) (Green, 1990). Indeed, the (strept)avidin-biotin interaction is considered one of the strongest non-covalent interactions known in nature (Kuntz et al., 1999). It is not well understood why living organisms would need a protein capable of binding biotin with such an extreme affinity. A role in sequestering excess biotin is unlikely, as even high amounts of biotin are non-toxic (Mock and Heird, 1997). Instead, it has been postulated that the secretion of streptavidin by bacteria represents a natural defence mechanism; the sequestration of any extracellular biotin by streptavidin will give the secreting bacteria an advantage over other biotin-requiring microorganisms in the environment. This is supported by streptavidin being present in the antibiotic complex MSD-235 in combination with stravidin, a low molecular weight compound which is able to inhibit the growth of Gram-negative bacteria (Baggaley et al., 1969).

1.1.4 Biotin analogues / derivatives

Numerous biotin analogues exist and are capable of binding to SA. However, all biotin analogues bind to SA with a reduced affinity compared to biotin, and hence have been used in biotechnology due to the less severe conditions needed to break the analogue-SA interaction (e.g. in the affinity purification of SA). The binding affinities of some of the biotin analogues / derivatives discussed below are given in Table 1.

Researchers have investigated whether it is possible to synthesise a biotin analogue that binds with higher affinity to SA or Av. For example, computational modelling has been utilised to assess how changes to biotin's structure might affect the binding free energy, resulting in the prediction that a biotin analogue containing a fluorine atom at the *pro-8R* position would have more favourable non-polar interactions with SA and thus increased binding affinity (Kuhn and Kollman, 2000). In a similar approach, molecular dynamic simulations of solvated biotin when free and when bound to SA and Av suggested that the 9*R*-methyl analogue of biotin (Figure 4) should bind more strongly to both SA and Av than biotin (Dixon et al., 2002). When 9*R*-methyl biotin was synthesised and its binding characterised, it was instead found that the binding affinity for both SA and Av was actually reduced, with considerably smaller exothermic enthalpy changes and a decrease in binding free energy of 1.3 kcal/mol for 9*R*-methyl biotin binding to SA. However, 9*R*-methyl biotin remains the strongest binding biotin analogue characterised so far.

The biotin derivative 2-iminobiotin (IB) contains a guanidine group in place of the ureido group in biotin, which results in the binding affinity being pH-dependent (Figure 4). At high pH (pH 9.5 or higher), the guanidine group is not protonated, and IB is able to bind SA with a high affinity ($K_d \sim 10^{-11}$ M) (Fudem-Goldin and Orr, 1990). At low pH ($\sim$ pH 4), the guanidine group becomes protonated, which dramatically decreases the binding affinity ($K_d > 10^{-3}$ M). This pH-dependency has been exploited in affinity chromatography, where IB immobilised on a solid support such as sepharose can be used to purify SA, with capture at high pH followed by elution at low pH (Hofmann et al., 1980).

Desthiobiotin (Figure 4) is a non-sulfur-containing biotin precursor, which binds SA with a moderate affinity ($K_d \sim 10^{-9}$ M) (Green, 1970). This easily reversible interaction can be used for the isolation and purification of desthiobiotin-labelled proteins (Hirsch et al., 2002). A *N*3'-ethyl biotin analogue, with a dissociation constant of 0.8 nM, has also been used in purification protocols to give easily reversible SA binding (Ying and Branchaud, 2011).

A large number of functionalised biotins have also been produced, by coupling via the carboxyl group on the valeryl chain of biotin (e.g. by 1-ethyl-3-[3-dimethylaminopropyl]carbodiimide hydrochloride (EDC) cross-linking to primary amines). These biotin derivatives are still recognised by anti-biotin antibodies, but their affinity for SA depends on the nature (i.e. composition and length) of the spacer between biotin and the linked functional group

(Hofmann et al., 1982). A commonly encountered biotin derivative is biocytin, formed by the linkage of biotin and lysine (Figure 4). Biocytin is often used as an intracellular labelling reagent for neurons (Thomson and Armstrong, 2011). Other biotin derivatives include: photoreactive biotin derivatives, such as *N*-(4-azido-2-nitrophenyl)-amino-propyl-*N'*-(*N*-D-biotinyl-3-aminopropyl)-*N'*-methyl-1,3-propanediamine, which are activated by strong visible or UV light and used to label DNA and proteins with biotin (McInnes et al., 1990); and fluorescent biotin derivatives, such as biotin-4-fluorescein (Kada et al., 1999), which can be used to identify and quantify biotin-binding sites (structure given in Section 3.2.3).

Biotin analogue	K _d (M) of:	
	SA binding	Av binding
Biotin	4 x 10 ⁻¹⁴ (Green, 1990)	6 x 10 ⁻¹⁶ (Green, 1990)
9 <i>R</i> -methyl biotin	3 x 10 ⁻¹³ (Dixon et al., 2002)	ND
2-iminobiotin	10 ⁻¹¹ (high pH); > 10 ⁻³ (low pH) (Fudem-Goldin and Orr, 1990)	10 ⁻¹¹ (high pH); > 10 ⁻³ (low pH) (Green, 1966)
Desthiobiotin	~ 10 ⁻⁹ (Green, 1990)	3.5 x 10 ⁻⁹ (Green, 1990)
<i>N</i> 3'-ethyl biotin	8 x 10 ⁻¹⁰ (Ying and Branchaud, 2011)	ND
Biotin-4-fluorescein	1.4 x 10 ⁻¹⁰ (Aslan et al., 2005)	ND

Table 1. Dissociation constants determined at 25 °C of a selection of biotin analogues for SA and Av binding. ND, not determined.

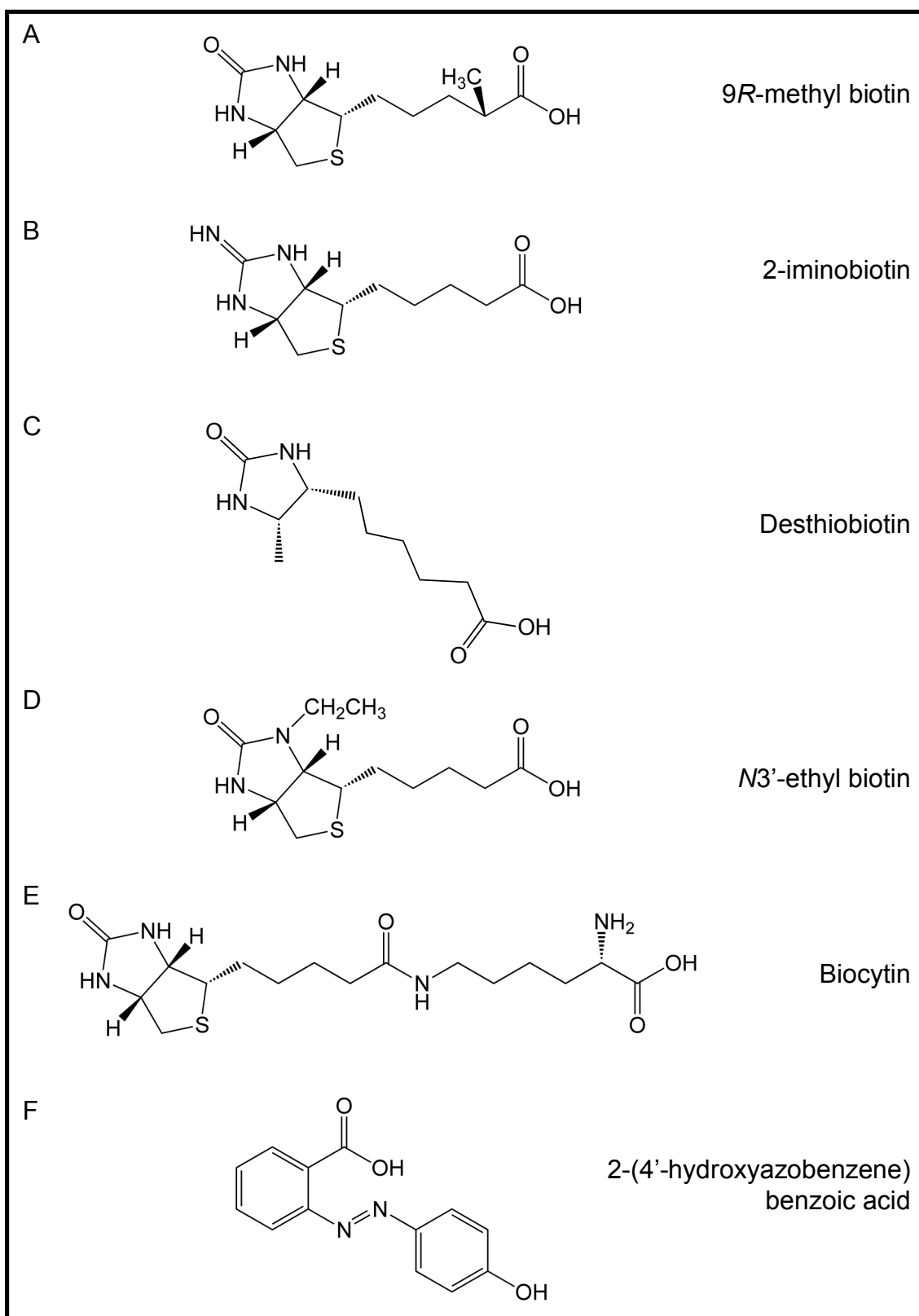


Figure 4. Examples of small molecules capable of binding streptavidin. A, 9*R*-methyl biotin; B, 2-iminobiotin; C, desthiobiotin; D, *N*3'-ethyl biotin; E, biocytin; F, 2-(4'-hydroxyazobenzene) benzoic acid (HABA) (see Section 1.1.5).

1.1.5 Other streptavidin binders

Not all SA binders closely resemble biotin in structure. 2-(4'-hydroxyazobenzene)benzoic acid (HABA) is a yellow dye that undergoes a red shift on binding to SA (Figure 4) (Janolino et al., 1996). Due to this colour change and its ease of displacement from SA by biotin, HABA can be used to quantify the biotin-binding capacity of SA and other biotin-binding proteins (Green, 1990). Crystal structures have revealed HABA binds in the biotin-binding pocket of SA in the same manner as biotin, with the benzoate ring of HABA making comparable interactions as the ureido group in biotin (Weber et al., 1992). HABA has a K_d of 10^{-4} M (Weber et al., 1992), but structure-guided engineering efforts have produced HABA derivatives with 165-fold increased binding affinity (Weber et al., 1994).

A number of peptide ligands able to bind to SA have also been developed, such as the Strep-tags and the Nano-tags. The well-known linear peptides Strep-tag (AWRHPQFGG) (Schmidt and Skerra, 1993) and Strep-tagII (WSHPQFEK) (Schmidt et al., 1996) have SA dissociation constants of 37 μ M (Schmidt et al., 1996) and 13 μ M (Voss and Skerra, 1997) respectively. These peptides are small, bind to the biotin binding-pocket in SA, can be genetically fused to proteins of interest without interfering with the protein's function, and can be released from the SA binding pocket by competition with biotin or desthiobiotin. The Nano-tags have slightly higher affinity for SA, with Nano-tag₉ (fMDVEAWLGAR) having a K_d of 17 nM and Nano-tag₁₅ (fMDVEAWLGARVPLVET) having a K_d of 4 nM (Lamla and Erdmann, 2004). Cyclisation of peptides via two cysteine residues that can disulfide bond

together results in a constrained conformation, which can give a dramatically increased SA binding affinity, in some cases increasing the affinity up to three orders of magnitude to the low micro-molar range (Giebel et al., 1995).

1.2 The avidin superfamily

1.2.1 Streptavidin structure

Streptavidin is a ~ 54 kDa protein, secreted by *Streptomyces avidinii* (Chalet and Wolf, 1964). The N and C termini of the full-length 159 aa protein are susceptible to proteolysis, and a heterogeneous mixture of truncated streptavidin molecules are produced in culture (Argarana et al., 1986). Commercial preparations of streptavidin utilise proteases to ensure complete truncation of the N and C termini, producing 'core' SA, consisting of residues 13 to 139 (Pahler et al., 1987). This 'core' streptavidin is the species most often used in biotechnology applications, as it is the minimum size of the protein that has the full biotin-binding activity whilst also having reduced tendency to aggregate (Pahler et al., 1987). The full-length protein has reduced biotin-binding ability, possibly due to residues at the C terminus folding back and occupying the biotin-binding site (Le Trong et al., 2006), as well as a higher tendency to aggregate (Sano et al., 1995).

The crystal structure of streptavidin was solved in 1989 (Hendrickson et al., 1989; Weber et al., 1989), followed in 1993 by the structure of avidin (Livnah et al., 1993; Pugliese et al., 1993). Both proteins are members of the (strept)avidin superfamily and are homotetrameric in structure. Each monomer consists of eight anti-parallel β -strands, that fold to give a β -barrel structure, with a biotin binding-site located at one end of the barrel. In this way, each streptavidin / avidin molecule is capable of binding up to four molecules of biotin, and is said to be tetravalent.

The tetrameric structure of SA can also be considered a dimer of dimers. A functional dimer is formed between adjacent subunits one and two (and three and four) (Livnah et al., 1993), as a result of each subunit contributing a conserved Trp-120 to the biotin binding-site of the other. The presence of this donated Trp-120 residue in the binding-site is essential for the high biotin affinity; when mutated to another amino acid, the biotin affinity decreases markedly (Chilkoti et al., 1995b; Laitinen et al., 1999), and the absence of this donated Trp-120 is thought to be a contributing factor to the decreased affinity always seen in monomeric SA (Qureshi and Wong, 2002; Wu and Wong, 2005). The most extensive subunit interface occurs between subunits one and four (and two and three), forming the structural dimer, with multiple van der Waals interactions, hydrophobic forces and hydrogen bonding contributing to the structural integrity of the tetramer.

1.2.2 The biotin binding site & the basis for high-affinity binding

The SA-biotin interaction is a truly remarkable non-covalent interaction. As a paradigm of a high-affinity protein-ligand system, the interaction has been the subject of extensive biophysical characterisation.

Firstly, the association (k_{on}) and dissociation (k_{off}) rates of the interaction have been determined. The extremely slow k_{off} has proven difficult to measure at 25 °C, and so is often experimentally determined at 37 °C and extrapolated back. Experiments utilising radiolabelled biotin revealed dissociation rates of $2.9 \times 10^{-5} \text{ s}^{-1}$ (Hyre et al., 2000) or $4.3 \times 10^{-5} \text{ s}^{-1}$ (Klumb et al., 1998) at 37 °C and $4.4 \times 10^{-6} \text{ s}^{-1}$ at 25 °C (Levy and Ellington, 2008). The k_{on} has been determined by a number of methods at 25 °C; stop-flow fluorescence experiments gave a k_{on} of $7.5 \times 10^7 \text{ M}^{-1}\text{s}^{-1}$ (Hyre et al., 2006) whilst surface plasmon resonance experiments gave a lower k_{on} of $5.1 \times 10^6 \text{ M}^{-1}\text{s}^{-1}$ (Qureshi et al., 2001). These association rates are slightly lower than what would be expected from diffusion-controlled binding ($\sim 10^8$ to $10^9 \text{ M}^{-1}\text{s}^{-1}$) (Berg and von Hippel, 1985), but can be explained by the total exposed surface area of the biotin-binding pockets comprising $\sim 2 \%$ of the total exposed surface area of SA (and hence not every collision resulting in binding) and the necessary displacement of five water molecules from the binding-pocket to allow binding (Qureshi et al., 2001).

Whenever SA features in a research paper, the remarkably low biotin dissociation constant of $4 \times 10^{-14} \text{ M}$, and sometimes 10^{-15} M (at 25 °C and pH 7) is quoted (Green, 1990). In the same paper, the dissociation constant for

K_d is given as 6×10^{-16} M (Green, 1990). However, the origins of the dissociation constant for SA are not immediately obvious, and on closer inspection, it was found that this dissociation constant was calculated using the off-rate for SA at pH 7 ($28 \times 10^{-7} \text{ s}^{-1}$) (Green, 1990) and the on-rate for Av at pH 5 ($7 \times 10^7 \text{ M}^{-1} \text{ s}^{-1}$) (Green, 1963), which they note has been measured only once. The dissociation constant for Av was calculated using the off-rate for Av at pH 7 ($0.4 \times 10^{-7} \text{ s}^{-1}$) (Green and Toms, 1973) and the on-rate for Av at pH 5 ($7 \times 10^7 \text{ M}^{-1} \text{ s}^{-1}$) (Green, 1963). A literature review reveals that a more appropriate figure to be quoting for the SA-biotin dissociation constant would be 5.9×10^{-14} M, arising from a biotin on-rate of $7.5 \times 10^7 \text{ M}^{-1} \text{ s}^{-1}$ measured by stop-flow fluorescence (Hyre et al., 2006) and an off-rate of $4.4 \times 10^{-6} \text{ s}^{-1}$ measured using radiolabelled biotin (Levy and Ellington, 2008), with both measurements being made at 25 °C and pH 7.

A range of techniques has also been used to investigate the thermodynamics of the interaction. Differential scanning calorimetry (DSC) has been used to investigate the effect of biotin-binding on the thermal stability of SA. Biotin-binding greatly increases the thermal stability of the tetramer; the denaturation temperature (at which half the SA molecules are denatured, T_m) is increased from 75 °C in apo-SA to 112 °C in biotin-bound SA (Gonzalez et al., 1999). DSC also revealed that the SA-biotin interaction is resistant to 6 M guanidine hydrochloride (GuHCl), with a T_m of 108 °C in the presence of this high concentration of GuHCl. Isothermal titration calorimetry (ITC) revealed an extremely negative free energy of biotin-binding of -18.3 kcal/mol, which has a large enthalpic component (-32 kcal/mol) (Weber et al., 1992). A survey of a

range of macromolecule-ligand interactions revealed a maximal free-energy contributions per nonhydrogen atom of ~ 1.5 kcal/mol, which decreased as the size of the ligand increased (Kuntz et al., 1999). The (strept)avidin-biotin interaction was a clear outlier from this trend (Figure 5).

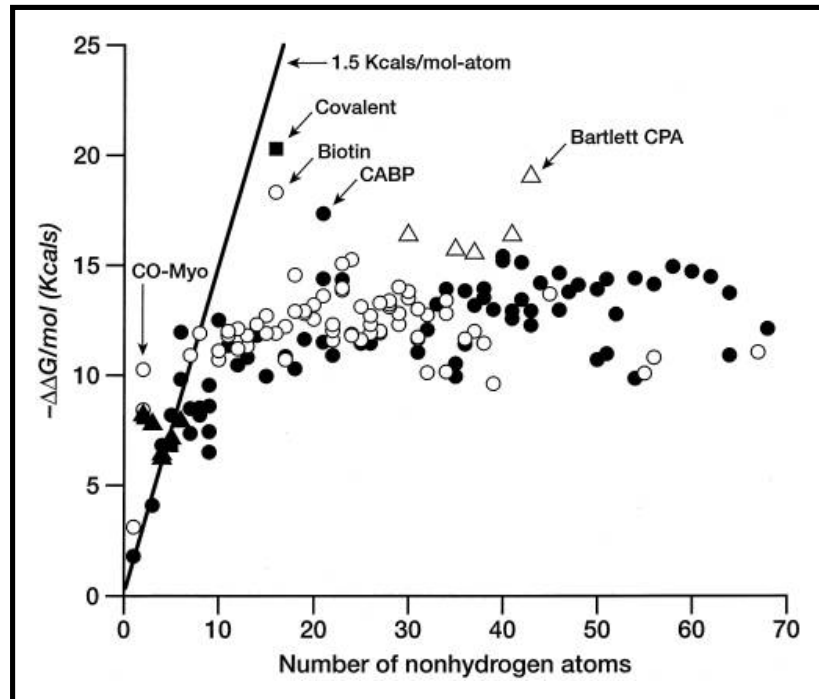


Figure 5. The free energy of binding for a range of macromolecule-ligand interaction (Kuntz et al., 1999). For smaller ligands, there is a maximal free energy of -1.5 kcal/mol per nonhydrogen atom, which decreases as the size of the ligand increases. The outlier corresponding to the avidin-biotin interaction is labelled (biotin), as are other outliers corresponding to carboxyl arabinitol bisphosphate (CABP) and Bartlett's carboxypeptidase (CPA). Copyright (1999) National Academy of Sciences, U.S.A. Permission obtained.

Since biotin-binding is remarkably favourable in enthalpy but adverse in entropy, it suggests biotin-binding increases internal bonding in SA whilst also restricting the protein's dynamic behaviour. Hydrogen/deuterium (H/D) exchange has clearly shown this restriction in SA flexibility on biotin-binding; in the absence of biotin, 72-74 backbone amide hydrogens in SA underwent exchange with solvent deuterium, compared to 48-50 exchanges when biotin

was bound (Williams et al., 2003). These ~ 24 backbone hydrogens that are protected from exchange reflect the reduced flexibility and increased stability of SA when biotin is bound. The region with the greatest reduction in exchange on biotin-binding is the loop connecting β -strands 3 and 3 (L3/4), reflecting the stabilisation of this flexible loop on biotin-binding (see below). Residues 71-78, which are found at the subunit interface in SA, undergo the least exchange in apo-SA, and experience very little further protection on biotin-binding, reflecting the restricted dynamics of this region. Despite biotin-binding improving the subunit packing and decreasing the flexibility of SA, biotin-binding is not positively co-operative, as the affinities of biotin-binding to SA containing zero, one, two and three biotin molecules are identical (Jones and Kurzban, 1995).

To this end, the crystal structures of SA and its mutants, which number over 100 in the Protein Data Bank (PDB), have helped shed light on the structure-function relationship of this interaction.

Firstly, it is immediately obvious when inspecting the structures of biotin-bound SA that the shape of the biotin-binding pocket in SA is highly complementary to the shape of biotin (Figure 6).

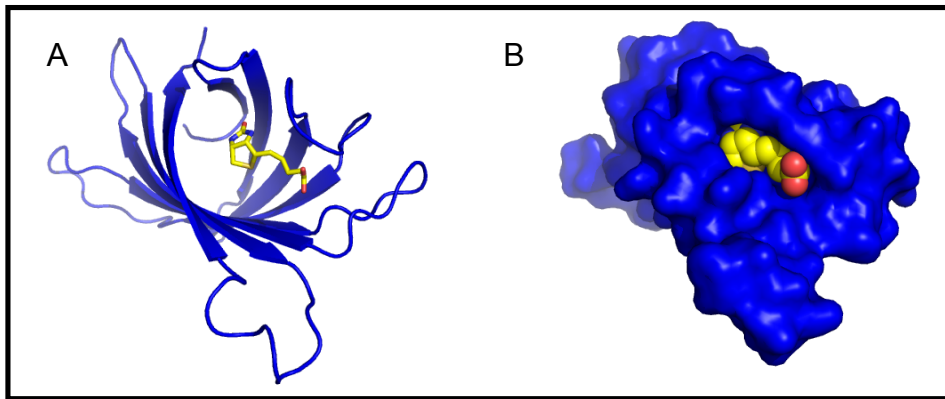


Figure 6. The biotin-binding pocket is highly complementary to the shape of biotin. One subunit of biotin-bound SA is shown in cartoon (A) and spacefill (B) representation (1SWE chain D).

Secondly, many contacts are made to biotin when it is in the binding pocket, involving van der Waals interactions with the multiple conserved tryptophans in the pocket (Trp-79, Trp-92, Trp-108 and Trp-120) (Figure 7), as well as an extensive network of hydrogen bonding. Miyamoto and Kollman argue for the majority of the negative free energy of biotin-binding arising from the van der Waals interactions (Miyamoto and Kollman, 1993), whilst Weber et al. argue for the importance of the hydrogen bonding network (Weber et al., 1992).

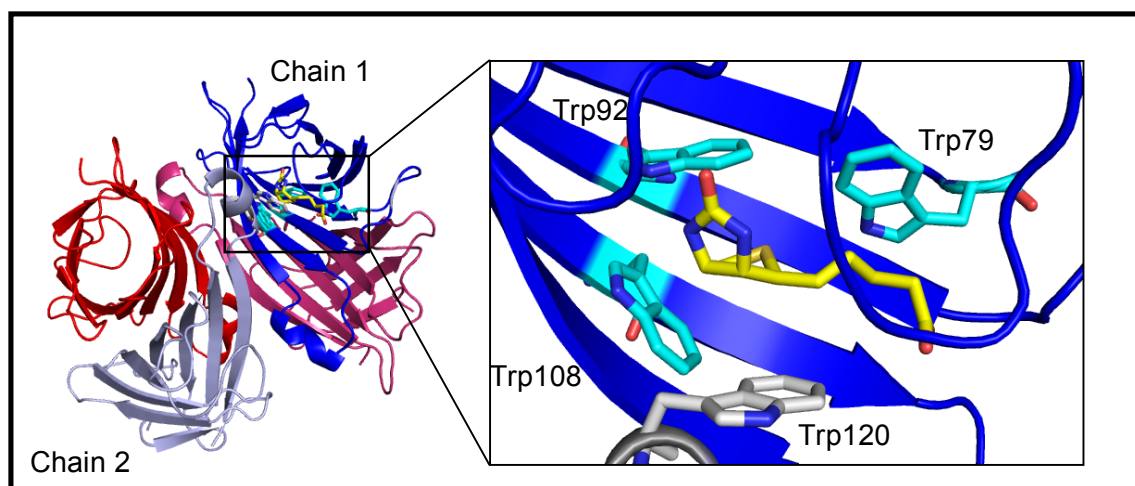


Figure 7. Biotin makes multiple van der Waals interactions with conserved tryptophans in the binding pocket. Tetrameric structure of SA, with only biotin bound to chain 1 shown. Close up view shows the biotin (carbon atoms coloured yellow) in the binding pocket of chain 1, with the conserved tryptophans shown in cyan sticks. The Trp-120 contributed by the other subunit in the functional dimer (chain A) is shown in grey. PDB code 1SWE (Freitag et al., 1997).

There are two shells of hydrogen-bonding residues in the binding pocket; a first shell consisting of eight residues that are directly hydrogen-bonded to biotin (Figure 8) and a second shell of residues that are hydrogen-bonded to the residues in the first shell. The ureido oxygen atom in biotin forms three hydrogen bonds to SA, which are arranged with tetrahedral geometry. In addition to this, the C-N bond lengths in biotin are shorter than in solution, whilst the C-O bond is longer than in solution, suggesting there is considerable polarisation of the ureido group when bound to SA (DeChancie and Houk, 2007; Weber et al., 1989).

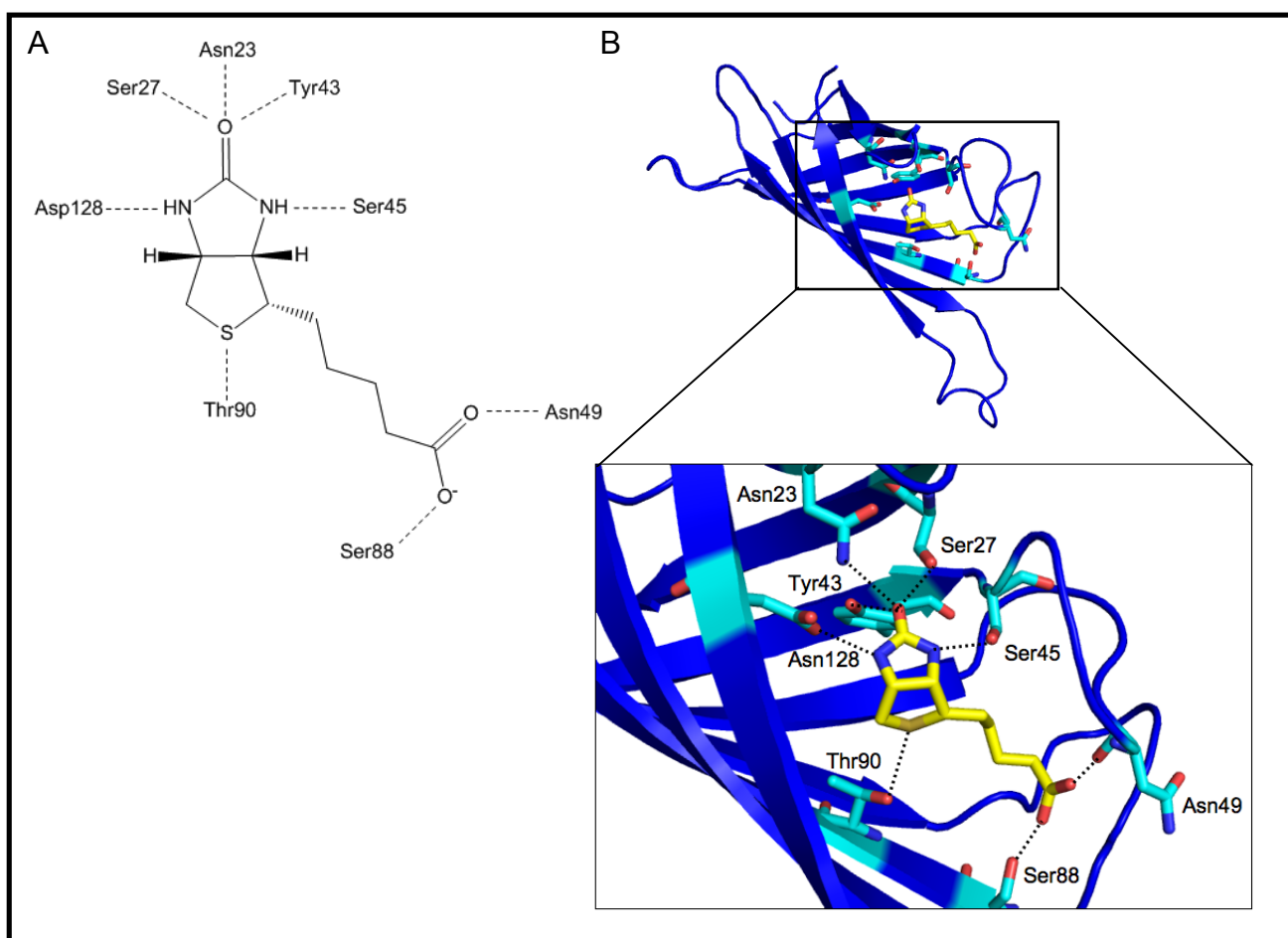


Figure 8. Hydrogen bonding network to biotin. A, Eight residues in SA form hydrogen bonds to biotin, contributing to the large negative free energy change on binding. B, Hydrogen-bonding residues in SA shown in cyan, with hydrogen bonds represented by black dashed lines. Biotin is shown in yellow stick form, with SA backbone (1SWE chain D) (Freitag et al., 1997) in blue cartoon.

Thirdly, biotin-binding is accompanied by the stabilisation of a flexible surface loop linking β -strands 3 and 4 (L3/4), which closes over the biotin binding site, forming a 'lid' over the pocket. L3/4 is disordered in apo-SA, and is very rarely seen in these structures. This ordering of L3/4, accompanied by quaternary structure changes causing tighter packing of SA subunits (Williams et al., 2003), results in biotin becoming 'buried' in the pocket, with only the valeryl carboxyl group accessible to the surrounding solvent. This stabilisation of L3/4, together with the capture of biotin in the pocket, would give unfavourable entropy effects, which could partly be balanced by the release of five water molecules that reside in the binding pocket in the absence of biotin (Chilkoti, 1995).

1.2.3 Other biotin-binders in the avidin superfamily

There are other members of the avidin superfamily that can bind biotin. Avidin-like proteins have been found in a range of species, including from chickens (Eakin et al., 1941; Laitinen et al., 2002) to bacteria (Chalet and Wolf, 1964; Helppolainen et al., 2007; Nordlund et al., 2005b), and from frogs (Maatta et al., 2009) to mushrooms (Takakura et al., 2009). The members of the avidin superfamily aligned in Figure 9, as well as those mentioned in the proceeding text, have the characteristic SA eight-stranded β -barrel subunit tertiary structure but bind biotin with varying affinities. The key properties of the members of the avidin superfamily discussed in the proceeding text are summarised in Table 2.

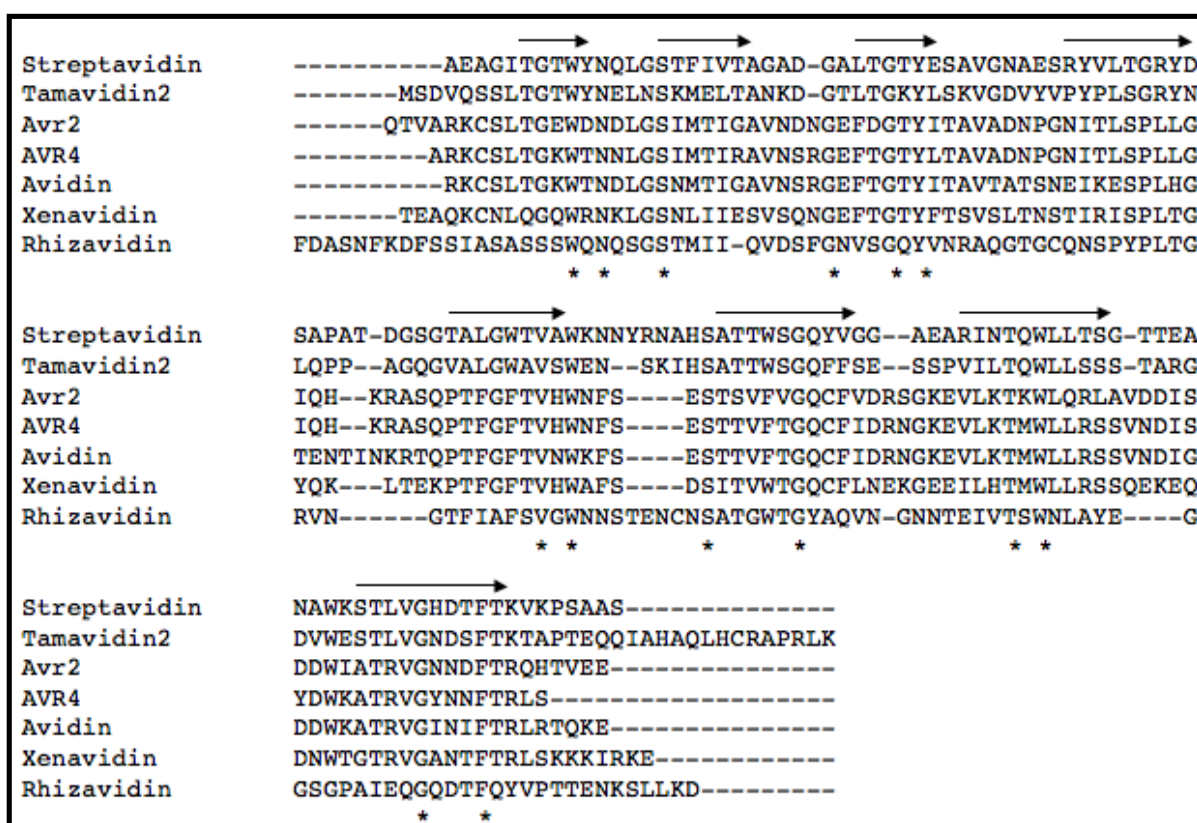


Figure 9. Multiple sequence alignment of members of the avidin superfamily. The pairwise alignment was carried out using ClustalW. The conserved residues are highlighted by an asterisk. The eight β -strands indicated by the arrows are positioned according to the structure of SA.

Protein	Molecular weight (kDa)	Quaternary structure at RT	Biotin K_d (M)	pI	T_m (°C)
Streptavidin	~ 54 ^a	Tetrameric	4×10^{-14} ^b	6.1 ^c	75 ^d
Avidin	63.1 ^c	Tetrameric	6×10^{-16} ^b	9.5 ^c	83 ^d
AVR1	ND	Tetrameric	4×10^{-8} ^e	7.3 ^e	83 ^f
AVR2	66.6 ^g	Tetrameric	5×10^{-8} ^e	4.7 ^e	91 ^g
AVR3	ND	Tetrameric	ND (irreversible) ^e	10.2 ^e	83 ^f
AVR4/5	42.1 ^g	Tetrameric	4×10^{-14} ^e	10.0 ^e	107 ^h
AVR6	51.5 ^g	Tetrameric	ND (irreversible) ^e	7.3 ^e	88 ^g
AVR7	ND	Tetrameric	ND (irreversible) ^e	7.3 ^e	83 ^f
Rhizavidin	29.0 ⁱ	Dimeric	ND	4.0 ⁱ	75 ⁱ
Bradavidin I	49.2 ^j	Tetrameric	ND	4.1 ^j	70 ^j
Bradavidin II	58.4 ^k	Tetrameric	ND	9.6 ^k	75 ^k
Burkavidin	~ 60 ^l	Tetrameric	$< 10^{-7}$ ^l	7.4 ^l	ND
Tamavidin 2	60.9 ^m	Tetrameric	ND	7.4 ^m	78 ^m
Xenavidin	74.5 ⁿ	Monomeric	ND	9.2 ⁿ	< RT ⁿ

Table 2. Properties of a selection of members of the avidin superfamily.

ND, not determined. T_m indicates the temperature of the mid-point of the thermally-induced transition from tetrameric apo-protein to monomeric protein (or from dimeric to monomeric protein in the case of rhizavidin). ^a Sano et al., 1995; ^b Green, 1990; ^c Nordlund et al., 2005b ^d Gonzalez et al., 1999; ^e Laitinen et al., 2002; ^f The AVR proteins were found to have comparable heat-stability to avidin, although the data were not shown (Laitinen et al., 2002); ^g Hytonen et al., 2005; ^h Hytonen et al., 2004; ⁱ Helppolainen et al., 2007; ^j Nordlund et al., 2005b; ^k Helppolainen et al., 2008; ^l Sardo et al., 2011; ^m Takakura et al., 2009; ⁿ The molecular weight stated is that of tetrameric xenavidin analysed by gel filtration (Maatta et al., 2009).

1.2.3.1 Avidin

Avidin (Av) was first discovered in chicken egg-whites, but has since been isolated from the eggs of other avian species, including duck, goose and ostrich (Hytonen et al., 2003). Avidin shares the same remarkable biotin-binding properties as SA, with a slightly higher affinity for free biotin (which arises from a lower off-rate (Green, 1990)) and similar resistance to heat (Donovan and Ross, 1973; Gonzalez et al., 1999), extremes of pH and denaturants (Green, 1975). The increased binding affinity is thought to arise from a greater number of interactions made to biotin when it is bound to Av (Figure 10); there are increased van der Waals interactions, due to the presence of two conserved tryptophans and two phenylalanines in the Av binding site (compared to just three tryptophans in SA), and a larger hydrogen bonding network (Av makes three additional H bonds to biotin than SA) (Livnah et al., 1993). However, Av is inferior to SA in its biotin-conjugate binding (Pazy et al., 2002) and has higher non-specific binding due to Av being glycosylated and having a pI of 9.5. Also, being a eukaryotic protein and containing a disulfide bond, Av is more difficult to express in bacteria. Many mutants of avidin have been made, with the most widely used and commercially available one being NeutrAvidin, a deglycosylated version of avidin with a near-neutral pI of 6.3, which has the same biotin-binding properties as the fully-glycosylated form of Av as well as low non-specific binding (Hiller et al., 1990).

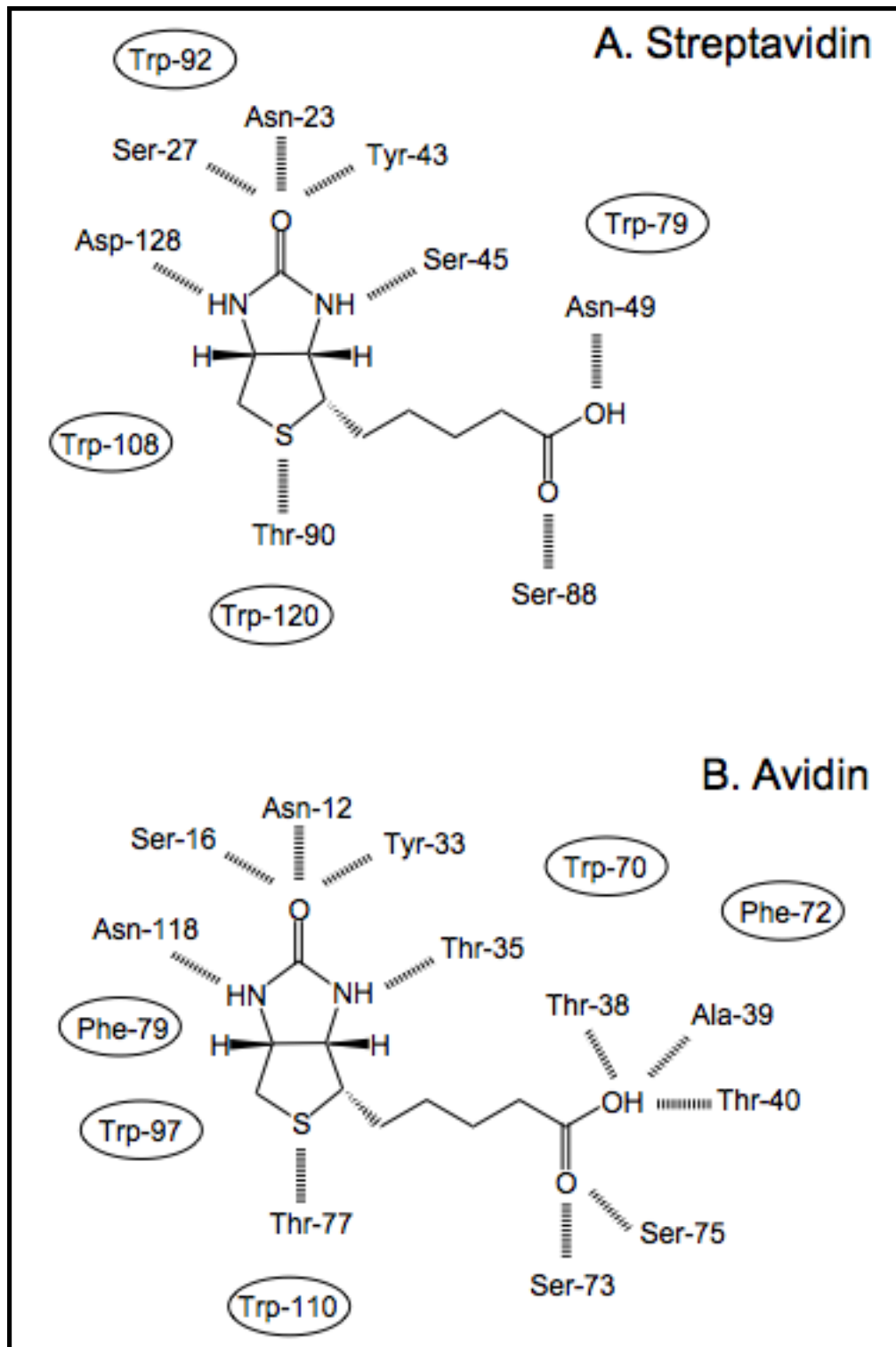


Figure 10. The increased biotin-binding affinity of Av is thought to arise from an increased number of interactions with the bound biotin compared to SA. A. The binding site in streptavidin consists of a hydrogen-bonding network involving eight conserved residues, as well as four conserved tryptophan residues (with Trp-120 being donated by a neighbouring subunit). B. In avidin, there are three conserved tryptophan residues (Trp-110 is donated from a neighbouring subunit), two conserved phenylalanine residues and eleven residues hydrogen-bonded to bound biotin.

Although Av and SA only share 30 % identity and 40 % similarity in their primary structure, their quaternary structures closely resemble one another. Worthy to note is avidin's longer L3/4, which, as in SA, is disordered in apo-Av but becomes ordered and stabilised when biotin binds (Livnah et al., 1993). However, L3/4 remains disordered when biotin-conjugates are bound, such as HABA, which could explain avidin's weaker conjugate binding when compared to SA (Repo et al., 2006). This longer L3/4 is also thought to impart Av with the ability to catalyse the hydrolysis of biotinyl *p*-nitrophenyl ester (BNP) (Figure 11), which SA is unable to do (Huberman et al., 2001). Also, Av contains a disulfide bond linking Cys-4 and Cys-83 in each subunit, which stabilises the tetramer in the absence of biotin, giving apo-Av a higher thermal stability (T_m of 83 °C) than apo-SA (Gonzalez et al., 1999). Indeed, mutagenesis leading to the removal of the disulfide bond from Av lowered the T_m to 76 °C, comparable to that of apo-SA (Nordlund et al., 2003).

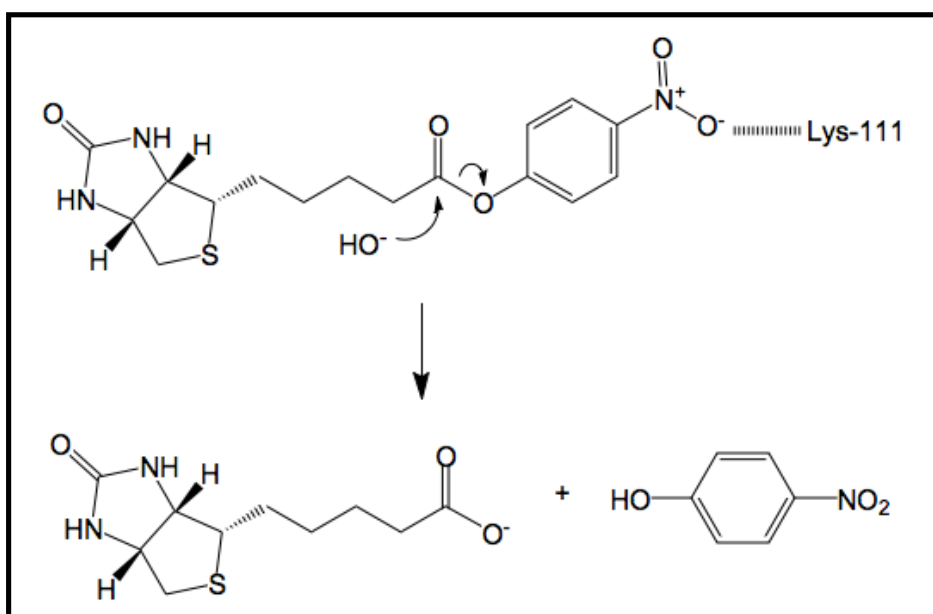


Figure 11. Av is able to catalyse the alkaline hydrolysis of biotinyl *p*-nitrophenyl ester (BNP). Lys-111 forms a hydrogen-bond to the nitro group of BNP, increasing the electron-withdrawing properties of the nitrophenyl group, thereby increasing the susceptibility of BNP to hydrolysis. Diagram adapted from Huberman et al., 2001.

1.2.3.2 Avidin-related proteins

By screening genomic libraries, seven avidin-related genes (AVR1 – AVR7) have also been found in chicken, that are 94 - 100 % identical to one another (AVR4 and AVR5 are identical and differ only by one nucleotide in their 5' flanking region) (Laitinen et al., 2002). However, the number of AVR genes varies between individual chickens, and even between cells within the same chicken, most likely due to unequal crossing-over and / or unequal sister chromatid exchange (Ahloth et al., 2001). The six AVR proteins (AVR1, AVR2, AVR3, AVR4/5, AVR6 and AVR7) have been expressed recombinantly and characterised, revealing that whilst the biotin-binding site residues are conserved, a number of amino acid replacements elsewhere in the proteins affect their physiochemical properties (Laitinen et al., 2002). For example, both AVR3 and AVR4/5 have a comparable basic pI to Av, whilst AVR1, AVR6 and AVR7 are neutral and AVR2 is acidic. However, all AVRs have comparable resistance to heat denaturation and proteolysis as Av, and have tetrameric quaternary structure. Considering biotin-binding, AVRs 3-7 have irreversible biotin-binding, as Av, whilst AVR1 has a K_d of $4.4 \pm 1.9 \times 10^{-8}$ M (18 % reversibility on addition of excess free biotin), and AVR2 has K_d of $5.2 \pm 1.7 \times 10^{-8}$ M (93 % reversibility on addition of excess free biotin) (Laitinen et al., 2002).

1.2.3.3 Avidin superfamily members isolated from bacteria

Rhizavidin was so named as the gene encoding this SA-like protein was found on a symbiotic plasmid in the nitrogen-fixing bacterial species *Rhizobium etli* (Helppolainen et al., 2007). Rhizavidin is an acidic protein (pI of

4.0), with comparable thermal stability to SA and high resistance to proteinase K (Helppolainen et al., 2007). Rhizavidin has a very similar tertiary structure to the other members of the avidin superfamily, consisting of eight antiparallel β -strands that fold into a β -barrel (Meir et al., 2009). However, it was the first member of the avidin superfamily to be identified as having a dimeric quaternary structure, rather than tetrameric. Furthermore, the loop connecting β -strands 7 and 8 (L7/8) is shorter and lacks the conserved Trp-120 that is usually donated to the biotin-binding site by the adjacent subunit in the functional dimer. Despite this lack of the conserved tryptophan, rhizavidin has strong biotin-binding affinity (Helppolainen et al., 2007), thought to arise from a unique disulfide bond linking L3/4 and L5/6, which could stabilise L3/4 and impede the dissociation of bound biotin (Meir et al., 2009). Rhizavidin is also a stronger biotin-conjugate binder than Av, although weaker than SA (there is 47 % release of the fluorescent biotin-conjugate ArcDia BF560 from rhizavidin after one hour at 50 °C, compared to 77 % release from Av and 22 % release from SA) (Helppolainen et al., 2007).

The genome of the nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* was also found to contain two putative genes for avidin-like proteins, which were named bradavidin I and II. Bradavidin I has comparable primary sequence similarity to SA as Av does, with 30.2 % identity and 41.7 % similarity (Nordlund et al., 2005b). Bradavidin I also has very similar high binding affinity towards biotin-conjugates, as SA does. However, unlike SA, bradavidin I has a lower affinity towards free biotin and contains two cysteine residues which form an intrasubunit disulfide bond linking L3/4 and L5/6.

Bradavidin II has low primary sequence similarity to bradavidin I (24 % identity and 36 % similarity) and an even lower similarity to SA (19 % identity and 32 % similarity) (Helppolainen et al., 2008). Also, like rhizavidin, bradavidin II lacks the conserved Trp-120 that is usually deemed essential for high-affinity binding and contains two cysteine residues that form an intrasubunit disulfide bridge between L3/4 and L5/6. Despite this, bradavidin II is still considered a strong biotin-binder, although it does bind both biotin and biotin-conjugates with weaker affinity than SA and bradavidin I (Helppolainen et al., 2008).

An avidin-like gene was also isolated from the human pathogen *Burkholderia pseudomallei*, named burkavidin (Sardo et al., 2011). This tetrameric biotin-binding protein is resistant to proteinase K digestion and possesses high thermostability, like SA. However, the cysteine residues that form an intra-monomeric disulfide bond in Av are conserved in burkavidin. ITC was used to monitor HABA binding to burkavidin, giving a K_d of $\sim 9 \times 10^{-8}$ M (Sardo et al., 2011).

1.2.3.4 Avidin superfamily members isolated from other species

The first avidin-like proteins found in organisms other than birds and bacteria were the tamavidins, isolated from the fungus *Pleurotus cornucopiae* (also known as the Tamogitake mushroom) (Takakura et al., 2009). A useful property of these biotin-binding proteins is that they are expressed as soluble proteins when recombinantly expressed in *E. coli*. Tamavidin 1 and 2 are tetrameric proteins with very high thermal stability, and share 47 % and 48 % sequence identity to SA respectively. Tamavidin 1 appears to be expressed

at a 20-fold higher level than tamavidin 2 in Tamogitake mushrooms, but despite binding biotin very strongly, does not bind 2-iminobiotin, which is an inconvenience when considering possible purification methods. Therefore, tamavidin 2 is thought to have the most potential as an alternative biotin-binding protein, with a K_d for 2-iminobiotin of $8.7 \pm 2.0 \times 10^{-8}$ M (Takakura et al., 2009).

The first avidin protein from the frog was discovered in 2009 in *Xenopus tropicalis* and so named xenavidin (Maatta et al., 2009). Xenavidin shares 56 % amino acid sequence identity with Av, and has a very similar quaternary structure, susceptibility to proteinase K digestion and pI. However, xenavidin has a much lower thermal stability in the absence of biotin, with the protein already dissociating into monomers at 22 °C, and has a higher off-rate from biotin and biotin-conjugates than avidin (Maatta et al., 2009).

1.3 Streptavidin engineering

There exists over 200 published mutants of SA (Laitinen et al., 2006; Laitinen et al., 2007). Introducing mutations into SA has altered: the binding-affinity towards biotin and other biotin-conjugates; the quaternary structure; and the thermostability of the protein. The range of protein characteristics observed on mutagenesis (as opposed to the complete abolishment of function and structure) suggests SA has a high degree of 'mutational robustness' and so should be a very good candidate for investigation by mutagenesis due to its high starting stability (Bloom et al., 2006).

A large variety of modification strategies have been utilised in SA engineering efforts. Chemical modification of SA has been used as a facile way of creating novel characteristics, for example, decreasing the kidney localisation of SA by reacting with succinic anhydride (succinylating amine groups) (Wilbur et al., 1998) or reacting with 1,2-cyclohexanedione (modifying arginine residues) (Wilbur et al., 2002). However, the major disadvantages of chemical modification are that it is difficult to achieve specific targeting of residues within the protein and difficult to obtain quantitative conversion. Most SA modifications, therefore, have used genetic engineering mutagenesis techniques, the two main methods being rational site-directed mutagenesis / design and directed evolution (Figure 12). The methods of rational design and directed evolution, and their application to SA engineering, are discussed below.

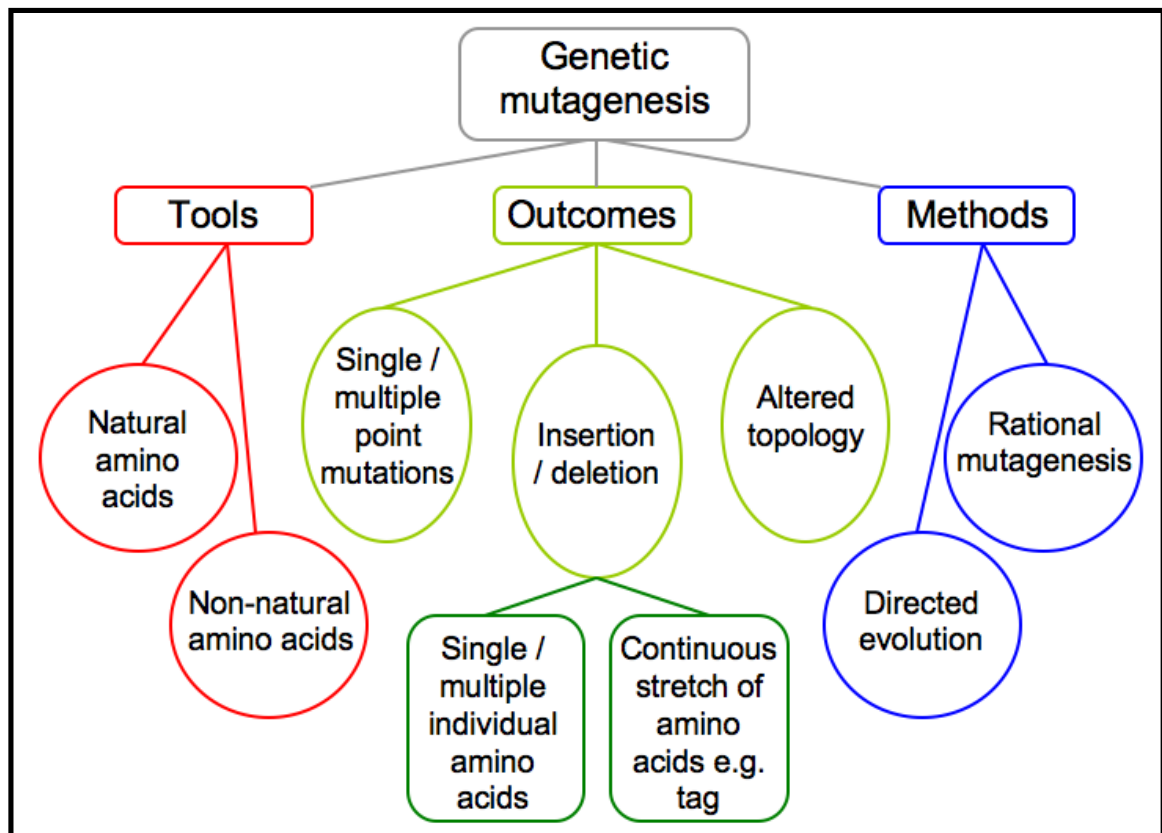


Figure 12. Genetic mutagenesis of proteins is an extremely powerful method of engineering new or altered functionalities.

1.3.1 Rational site-directed mutagenesis

A rational site-directed mutagenesis approach to protein engineering involves mutating, inserting or deleting specific residues in a protein in order to achieve a desired characteristic (Cedrone et al., 2000) (Figure 13). This approach usually makes use of the structure of the starting protein to guide the choice of residues to be mutated. For example, if a protein's binding specificity or affinity is to be altered, the rational choice would be to mutate residues in and around the binding site. However, identification of the conserved residues within a protein family can also give a guide to which are necessary for function. In addition, the impact of mutagenesis can be predicted to some degree *in silico* using computer simulations and modelling (Offman et al., 2011).

The techniques utilised in rational mutagenesis are well understood, facile and inexpensive, an advantage over the technically challenging and sometimes expensive protocols used in directed evolution. Rational mutagenesis also allows testing of hypotheses of how a protein's function arises from its primary structure.

A major disadvantage to rational site-directed mutagenesis is that it is time- and labour-intensive, and so does not enable screening of anywhere near the large numbers of mutants that can be created and screened using a directed evolution approach. Also, knowing the structure of the protein being mutated and having some knowledge of the protein's structure-function relationship is very helpful, but not essential. A working hypothesis of how the desired

characteristics might be introduced / altered in the protein is needed prior to selection of mutations. Therefore, the success of rational mutagenesis can be largely dependent on how well the effect of making certain changes to the protein can be predicted.

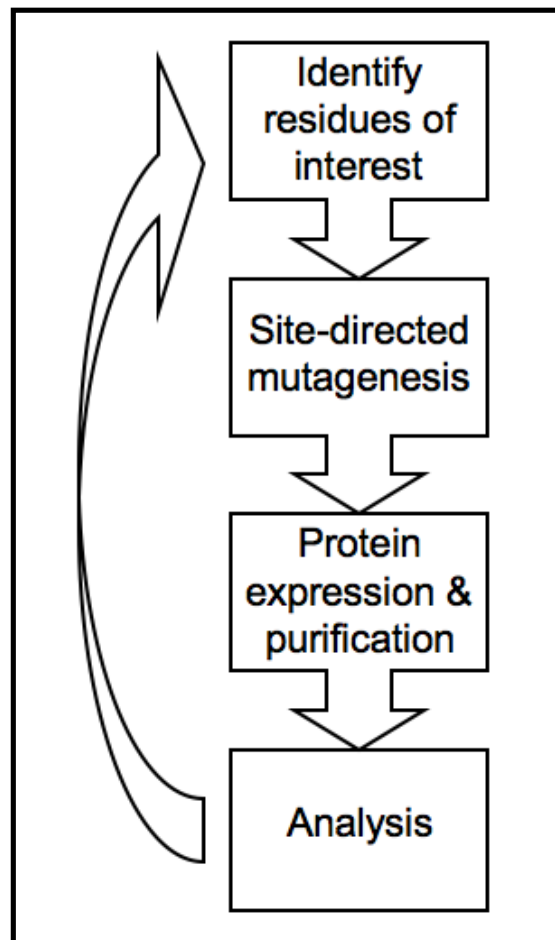


Figure 13. Flow diagram of the process of rational mutagenesis. Residues of interest in the starting protein can be identified by inspecting the structure of the protein or identifying conserved residues by sequence alignments. These residues are then mutated, and the resulting protein expressed, purified and analysed for the desired characteristic. The whole process can be repeated as many times as needed, introducing new mutations in each round.

1.3.1.1 Rational site-directed mutagenesis of SA

In order to gain insights into the origins of the high biotin-binding affinity of SA, the residues in and around the binding site have been major targets of rational site-directed mutagenesis. All of the mutations in this vicinity resulted in a lowered biotin affinity, with the majority of the mutations resulting in a greatly decreased affinity (Wilchek and Bayer, 1999). The importance of the conserved tryptophans in the binding pocket was highlighted when mutation of each to alanine decreased the biotin-binding affinity greatly (Chilkoti et al., 1995b). In particular, the conserved Trp-120 residue, which is donated to the binding site of one subunit by a neighbouring subunit, has been the subject of extensive mutagenesis, and was found to be important for generating both the high biotin-binding affinity and also the high tetramer stability of SA (Sano and Cantor, 1995).

The importance of the hydrogen-bonding network in generating high-affinity binding was also demonstrated using site-directed mutagenesis. Seven of the eight residues in the first shell of the hydrogen-bonding network (i.e. they form hydrogen bonds directly to the bound biotin; see Figure 10) have been mutated, both individually and in combination with other mutations. The three residues hydrogen-bonded to the ureido oxygen in biotin (Asn-23, Ser-27 and Tyr-43) have each been individually mutated to alanine, which gave a decreased biotin-binding affinity in all three mutant proteins produced (Klumb et al., 1998). Furthermore, mutations to Asn-23 and Ser-27 have been combined with other mutations to give mutant SAs with dramatically decreased affinity; the double mutant N23A S27D, with a K_d of 1×10^{-4} M

(Reznik et al., 1998), was combined with another hydrogen-bonding mutant, S45A, to give a triple SA mutant with negligible biotin-binding ability (Howarth et al., 2006). This N23A S27D S45A mutant of SA is often called 'dead SA' and has found considerable use in many applications that utilise controlled valency SA (Blois et al., 2009; Hong et al., 2010; Huang et al., 2010). The residues hydrogen-bonded to the ureido nitrogens (Ser-45 and Asp-128) and the sulfur atom in the tetrahydrothiophene ring (Thr-90) of biotin have also been the subjects of site-directed mutagenesis (Qureshi et al., 2001). Mutation of these residues was found not only to impact the biotin-binding affinity (the K_d of the mutants resulting from single mutations of these residues to alanine were in the range 10^{-9} to 10^{-11} M), but also the stability of the tetramer. The triple mutant S45A T90A D128A had monomeric quaternary structure and a K_d of 2×10^{-6} M.

Other site-directed mutagenesis efforts have focussed on the residues at the interfaces between the subunits within the SA tetramer. Attempts to stabilise the SA tetramer by increasing the electrostatic interactions across the dimer-dimer interface and by introducing intersubunit disulfide crosslinks have increased the thermal stability of SA with little effect on the biotin-binding ability of the tetramer (Reznik et al., 1996). SA mutants with lower valencies have also been produced by introducing destabilising mutations at the tetramer interfaces. Mutation of the conserved Trp-120, which is donated to the biotin-binding pocket of one subunit by the neighbouring subunit, reduces the binding affinity by several orders of magnitude (K_d of the W120K mutant $\sim 1 \times 10^{-8}$ M), as well as destabilising the tetrameric structure to give dimeric SA

(Laitinen et al., 1999). Other destabilising mutations have been introduced at the interfaces to give monomeric SA mutants, but all monomeric SAs reported to date have dramatically lowered biotin-binding affinity, due to their incomplete binding-pocket (due to the absence of the critical Trp-120 residue). The mutations T76R and V125R introduced sufficient electrostatic repulsion and steric hindrance at the SA tetramer interface to give monomeric SA, with a K_d of 1.7×10^{-7} M (Wu and Wong, 2005). To minimise aggregation of this mutant, two further mutations were introduced to increase the hydrophilicity of the resulting monomer, V55T and L109T. The resulting quadruple-mutant monomer had a K_d of 1.9×10^{-7} M and can be used as an affinity matrix, giving reversible capture of biotinylated molecules (Wu and Wong, 2006).

Rational mutagenesis has also been used to design topological mutants of SA. The role of the loop connecting β -strands 3 and 4 (L3/4) in generating the high biotin-binding affinity was investigated by the creation of a circularly permuted SA, in which the N and C termini were linked together and new termini introduced in L3/4 (Chu et al., 1998). This SA mutant possessed the characteristic SA quaternary structure of a tetramer of β -barrels, but the biotin affinity was dramatically reduced ($K_d \sim 10^{-7}$ M). An alternative topological SA mutant was created by the fusion of two different circularly permuted SA subunits into a single chain, followed by directed evolution using error-prone PCR and selection with biotinylated beads (Aslan et al., 2005). This mutagenesis strategy resulted in the production of dimeric and tetrameric SA mutants with decreased affinities for biotin, but increased relative affinities

towards the fluorescent biotin conjugate, biotin-4-fluorescein (structure given in Chapter 3 Methods).

Other properties of SA have been altered by site-directed mutagenesis. The non-specific cell binding properties of SA were reported to be decreased by site-directed mutation of the RYDS motif (residues 59 to 62), which was selected for mutagenesis due to its similarity to the RGDS integrin-binding motif (Murray et al., 2002). The antigenicity of SA was also decreased by identifying solvent-exposed aromatic and charged residues in SA and conducting site-directed mutagenesis at these positions (Meyer et al., 2001). Although this approach resulted in the production of a large number of mutants for characterisation, it did result in producing SA mutants that had both a decreased antigenicity and also a decreased biotin off-rate. This decreased biotin off-rate was extremely surprising, although very little attention was drawn to this fact in the paper (perhaps as it was only measured once for each mutant) and the authors offered no explanation of its origins.

The ease with which SA can be conjugated to other molecules has been increased by both site-directed mutagenesis of a single residue and also the insertion of short substrate sequences into SA. To enable conjugation of SA by covalent linkage to a thiol-, maleimide- or iodoacetamide-containing molecule, a cysteine residue was introduced into SA (Chilkoti et al., 1995a). In addition, a six or seven glutamine-containing amino acid sequence has been inserted into the SA sequence to enable conjugation to amines by transglutaminase catalysis (Lin and Ting, 2006), and a pentapeptide motif has

also been used to give conjugation to glycine-tagged molecules by sortase / transpeptidase catalysis (Matsumoto et al., 2011).

Non-natural amino acid mutagenesis is another facet of rational site-directed mutagenesis (Bain et al., 1989; Noren et al., 1989). As with an increasing number of proteins, it has been shown that incorporating non-natural amino acids at desired positions in SA is possible, which both increases the variety of modifications that can be made to SA and expands the spectrum of potential SA characteristics. Fluorescent dansylated non-natural amino acids (Hohsaka et al., 2004), fluorescent anthraniloyl-containing non-natural amino acids (Taki et al., 2001) and modified lysine residues (Tokuda et al., 2011) have all been introduced into SA using an *in-vitro* translation system.

1.3.2 Combining rational mutagenesis with directed evolution

Directed evolution is the deliberate evolution of macromolecules towards new functionalities (Jackel et al., 2008). It involves iterative cycles of mutagenesis, to produce a library of mutants, followed by selection and amplification of library members with more favourable characteristics, such as an increased catalytic ability. An overview of the process, both *in vitro* and *in vivo*, is shown in Figure 14.

Using a directed evolution approach to protein engineering has a number of advantages over rational design. A considerable advantage is that no prior knowledge or hypothesis of how the desired characteristic might arise is needed, whereas this is essential in rational design. Indeed, directed evolution outcomes can be surprising, when a protein with a desired characteristic is found to contain unexpected mutations (el Hawrani et al., 1996). Directed evolution can also be much more efficient than rational design, as it allows the simultaneous evaluation of a large number of mutants, which enables many more mutants to be tested in a short time period than is possible with rational design (Arnold, 1998).

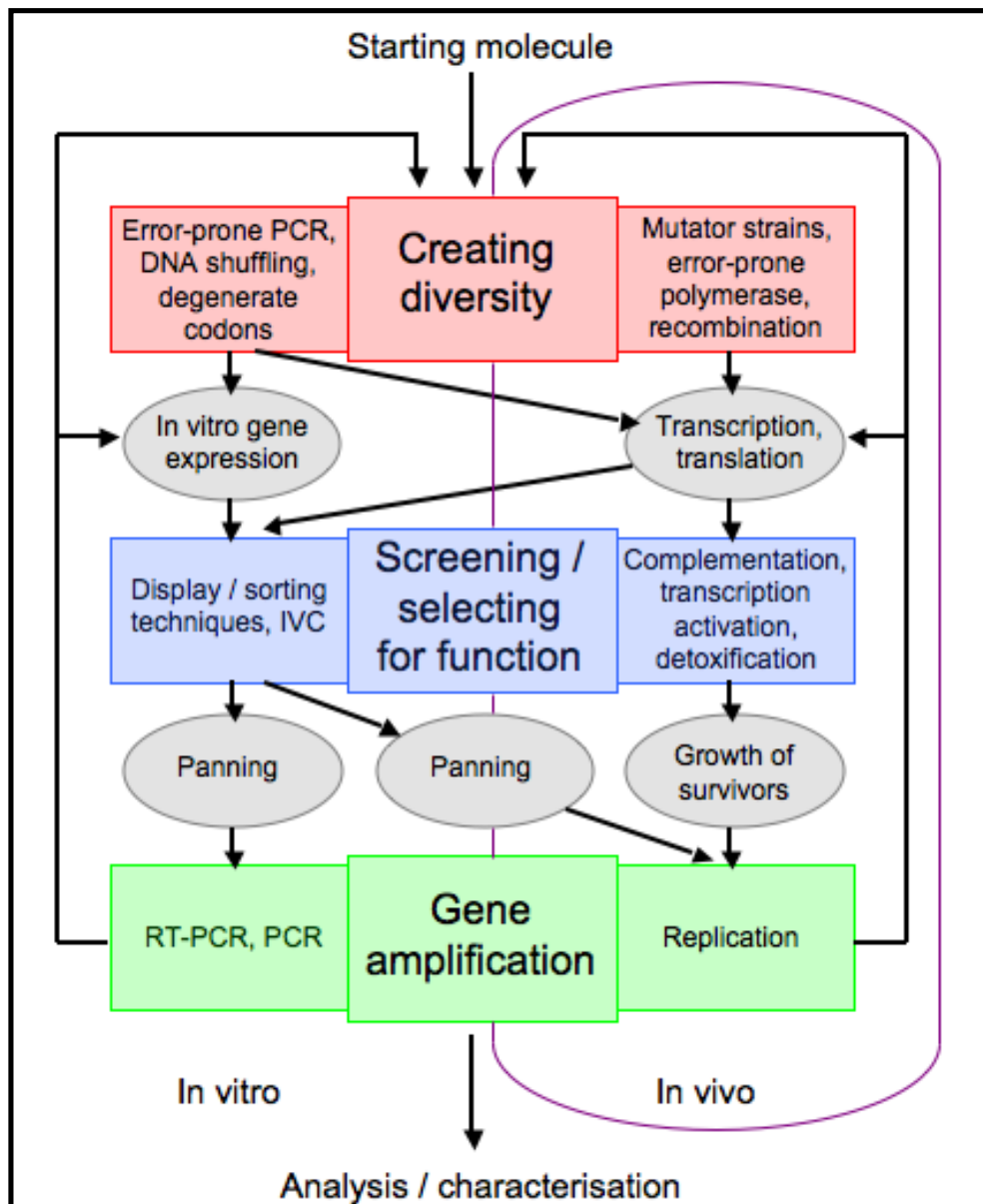


Figure 14. Directed evolution utilises a number of techniques in the iterative process of creating diversity, screening for desirable characteristics and then amplifying the selected mutants. Figure adapted from Jackel et al., 2008.

Powerful, high-throughput selection and screening methods for directed evolution exist, and can even be automated (Esvelt et al., 2011), but the protocols used in these steps can be technically difficult and involve expensive robotics. Also, not all desired characteristics can be easily screened for, such as inconspicuous target molecules that can not be optically detected (Dietrich et al., 2010). ‘You get what you screen for’ is commonly

regarded as the First Law of Directed Evolution (Schmidt-Dannert and Arnold, 1999), stressing the importance of the screening substrate being identical, or as close as possible, to the target substrate. Other factors in addition to the selection and screening methods can also influence the success of a directed evolution approach, including the design and quality of the mutant library and the methods used for creating and screening for the diversity in the library (Dietrich et al., 2010). However, overall, directed evolution remains a rapid and efficient method of protein engineering.

Although a large amount of SA mutagenesis has been carried out using a rational design approach, rational engineering and directed evolution have been used in tandem to give a powerful protein engineering strategy. In this way, rational design can be used to define or focus the mutant library diversity in directed evolution approaches. For example, rational design can be used to introduce an initial level of a desired characteristic into the starting protein, which directed evolution can then rapidly expand and enhance, leading very quickly to an optimised level of activity (Yuan et al., 2005a). This tandem, combinatorial approach of using rational design followed by directed evolution has been used in SA engineering efforts as described below.

1.3.2.1 A tandem rational mutagenesis / directed evolution approach to SA engineering

Rational mutagenesis and directed evolution both have advantages over each other. As they can be considered complementary mutagenesis techniques, they are often used in tandem to give a powerful combinatorial mutagenesis strategy. In this way, directed evolution in combination with rational design has been used as a method for engineering SA mutants with altered functionality. As mentioned in Section 1.3.1.1, after rational design was used to create a novel single-chain SA dimer, a directed evolution strategy was adopted to screen for, and to increase, the binding affinity of the mutants towards biotin (Aslan et al., 2005). Although the selection step was based on affinity for biotinylated beads, the final mutants isolated by this strategy had a surprising 10^4 - 10^5 –fold preference for the fluorescent conjugate biotin-4-fluorescein (B4F) (structure given in Chapter 3 Methods), in comparison to wildtype SA, which has a 10^4 -fold preference for biotin over B4F. Analysis of the mutations that accumulate in the mutants after three rounds of mutagenesis by error-prone PCR revealed that most were at the surface of the protein and were very rarely at or near the binding site or subunit interfaces.

A combinatorial alanine-screening method (Weiss et al., 2000) has also been used to investigate the role of 38 C-terminal residues in SA (Avrantinis et al., 2002). A phage-display library was created in which these residues were allowed to vary only between the wildtype and alanine residues, with the resulting mutants selected for by biotin-binding and then sequenced. The

alanine : wildtype ratio at each position was determined as a means of measuring the importance of the wildtype side-chain at that position (i.e. a low ratio indicated the wildtype residue was preferred over alanine at that position for SA function). This approach revealed the importance of side-chain residues at a distance from the binding-site for SA's biotin-binding function, such as Tyr-96, Arg-103 and Phe-130.

A combined rational design / directed evolution approach was also used to engineer SA mutants with decreased dissociation from the biotin analogue desthiobiotin (Levy and Ellington, 2008). Rational design based on the crystal structure of SA was used to identify five residues in the biotin-binding pocket (Trp-79, Thr-90, Trp-92, Trp-108 and Leu-110) that were then randomised to create libraries that were screened for the ability to bind desthiobiotin. After four iterative rounds of this randomisation followed by screening, the libraries were further diversified by mutagenic PCR and DNA shuffling and then subjected to another three rounds of screening, selection and mutagenesis. At the end of the process, mutations had accumulated in the five initial randomised residues and also in other regions of SA. The resulting mutants had a variety of characteristics, including increased solubility when expressed in *E. coli*, increased affinity towards desthiobiotin as desired, unchanged affinity towards free biotin and a decreased affinity towards free biotin.

1.4 Applications of streptavidin

SA is an extremely widely used tool in many different biotechnological applications. It is used as: a molecular anchor on biosensor chips and arrays; a building block in nanoassembly; a detector of biotinylated molecules, when fused to a reporter molecule such as a fluorescent dye, quantum dot (QD) or horseradish peroxidase (HRP); and a carrier of radioactivity in Targeted Radionuclide Therapy to name only a few applications. Some of these applications are discussed in more detail below. Controlled-valency SA (such as monovalent and divalent SA) (Howarth et al., 2006) has also increased the diversity of SA applications. Divalent SA has been used in immobilisation protocols to avoid the multimerisation possible with tetrameric SA, which is important when measuring adhesion forces (Huang et al., 2010), whilst monovalent SA has been utilised as a steric trap in a novel method of measuring the dissociation constant of high-affinity protein interactions (Hong et al., 2010) and to drive unfolding of a target protein (Blois et al., 2009).

There are a huge number of commercially available products on the market that are based on, or involve, streptavidin, such as: dye-labelled SA; nanoparticle-conjugated SA (including QDs, magnetic beads and agarose beads); and kits for the detection and isolation of biotinylated proteins, DNA, RNA and their binding partners. The sheer number of research and development tools and techniques that utilise the SA-biotin interaction gives a good indication of the tremendous value of SA.

1.4.1 Targeted radionuclide therapy

Traditional Targeted Radionuclide Therapy (TRT) utilises a radionuclide-carrying antibody that recognises some tumour-cell specific antigen (Figure 15). However, this approach has a low tumour / background ratio and a slow tumour accumulation rate and so modern techniques have included additional steps in the protocols, usually involving the separate administration of the targeting molecule followed by the therapeutic molecule (Nordlund et al., 2005b). The (strept)avidin-biotin interaction is utilised in both two-step and three-step TRT. One example of two-step TRT involves the administration of a biotinylated monoclonal antibody targeted to cancer cells, followed some time later by the administration of radiolabelled SA / Av. This has the advantages of giving an extended time for the biotinylated antibody to bind, optimising tumour accumulation, before the administration of radiolabelled SA / Av. To prevent endogenous biotin blocking the binding of the radiolabelled SA / Av, and also to 'mop up' any freely circulating biotinylated antibody, a three-step TRT approach can be used, where a clearing agent such as unlabelled Av (which has a faster clearance time than SA) is administered after the biotinylated antibody, and before the administration of radiolabelled SA / Av (Boerman et al., 2003). An alternative three-step TRT protocol involves the use of a streptavidin-conjugated monoclonal antibody to target tumour cells, followed by the administration of a synthetic clearing agent, finally followed by the addition of radiolabelled biotin.

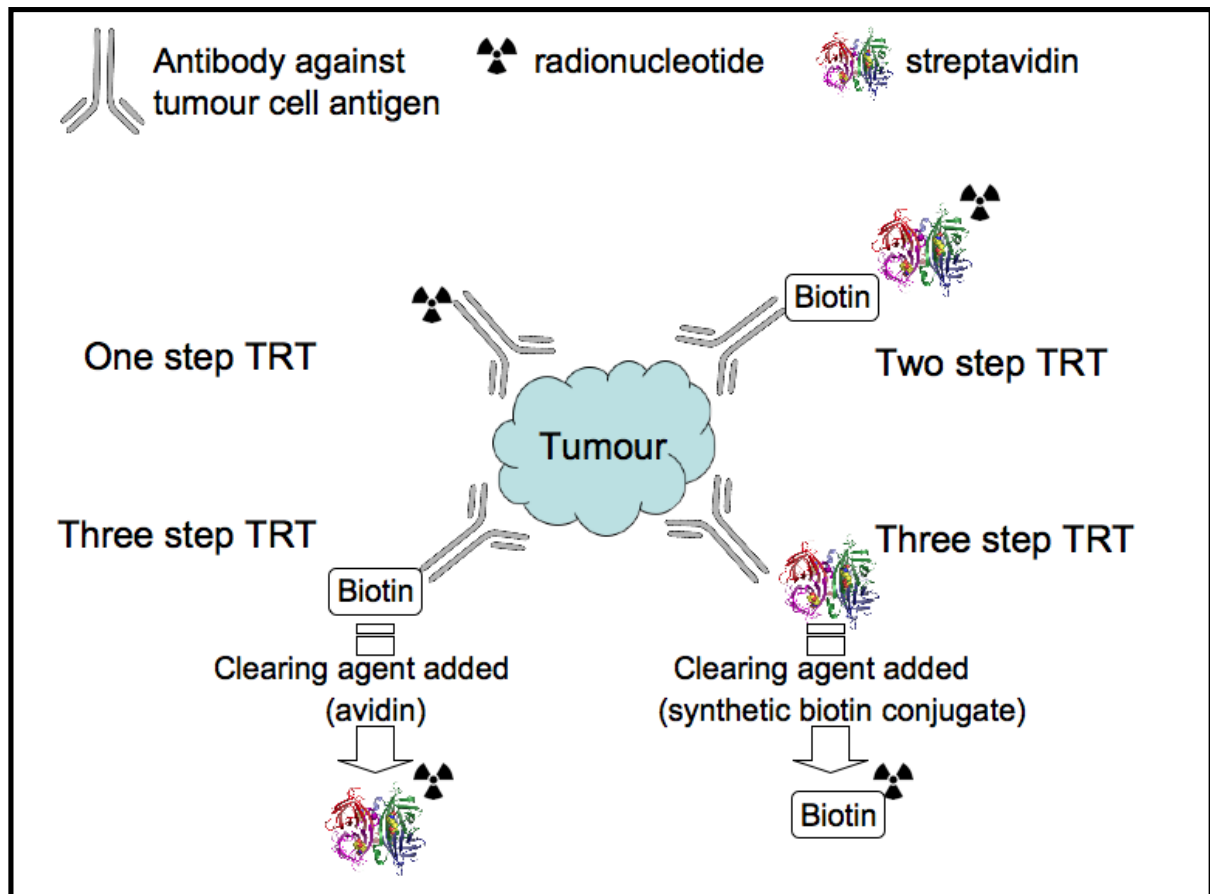


Figure 15. Targeted Radionuclide Therapy (TRT) can utilise a one-step, two-step or three-step protocol to deliver radioactivity directly to the tumour cell.

Repeated administration of SA / Av in humans can trigger an immunological response, and it has been proposed that other members of the avidin superfamily (e.g. the AVR proteins) could be used for successive treatments to avoid triggering a strong immunological response (Laitinen et al., 2002). However, recent research has suggested that avidin's immunogenicity does not affect its safety and efficacy in human cancer trials (Petronzelli et al., 2010). Several research groups are utilising the (strept)avidin-biotin system in their clinical trials of therapies against various cancers, including adenocarcinoma (Breitz et al., 2000), high-grade glioma (Grana et al., 2002) and non-Hodgkin's lymphoma (Weiden et al., 2000).

1.4.2 Cellular imaging

The SA-biotin interaction has also found use in cellular imaging. Alternative cellular imaging techniques include immunofluorescence microscopy and genetically-encoded fluorescent tags. In immunofluorescence microscopy, labelled antibodies raised against an antigen are used to specifically visualise a protein of interest. The antibodies are often labelled with either a fluorescent dye, such as the Alexa Fluor range of dyes (Invitrogen), or a QD, which is a semiconductor nanocrystal that is very bright and has a narrow emission spectra and negligible photobleaching (Barroso, 2011). However, in comparison to fluorescently-labelled antibodies, SA is dramatically smaller in size than an antibody (56 kDa for SA compared to 150 kDa for an average IgG antibody), antibody-antigen interactions span a range of affinities (see Section 1.5.3), most of which do not approach the extremely high affinity of the SA-biotin interaction, and certain antibodies can suffer from high levels of non-specific binding, giving high background signals. The use of genetically-encodable fluorescent proteins (FPs), based on green fluorescent protein (GFP) and its variants (Day and Davidson, 2009; Tsien, 1998) avoids the problems of specificity as the fluorescent reporter protein is covalently linked to the protein of interest, enabling direct visualisation of the protein in living cells. However, the large size of FPs (over 200 amino acids) can perturb the normal functioning of the protein of interest (Andresen et al., 2004) and many FPs are susceptible to bleaching over long imaging times (Seward and Bagshaw, 2009).

The SA-biotin interaction is used in cellular imaging applications due to the high specificity, speed of formation and stability of the interaction. Furthermore, SA can be easily conjugated to FPs, dyes and QDs, and gives low non-specific binding. Membrane-protein trafficking can be studied by using non-specific biotinylation of cell-surface proteins (often using an amine-reactive biotin) (Hurley et al., 1985) followed by visualisation with a fluorescently-labelled SA. However, this method results in biotinylation and visualisation of all surface proteins rather than a specific protein of interest, and also gives no control over where in the protein the amine-reactive biotin will bind, which can perturb normal protein function. To overcome this and obtain biotinylation of a protein of interest at a specific location in the protein, the BirA / AP system can be used (Chen et al., 2005; Howarth et al., 2005).

1.4.2.1 The BirA / AP system

The BirA / AP system exploits the ability of the *E.coli* enzyme biotin ligase (BirA) to biotinylate a certain lysine in a specific 15 amino acid sequence, called the acceptor peptide (AP) (Beckett et al., 1999). This AP sequence (GLNDIFEAQKIEWHE), also known as the AviTag™, can be genetically fused to the protein of interest, giving site-specific biotinylation of the protein of interest (Figure 16). In cells that have been transfected with DNA encoding the BirA protein fused to a signal peptide, BirA is expressed, folds in the endoplasmic reticulum and specifically biotinylates only proteins containing an AP tag; endogenous mammalian proteins are not biotinylated by BirA (de Boer et al., 2003). Alternatively, if the AP tag is attached to a protein that is directed to the cell surface, BirA can be expressed

recombinantly and then added to the extracellular medium, together with biotin, to give biotinylated cell surface proteins (Howarth et al., 2005). The biotinylated protein of interest can then be visualised by labelling with SA conjugated to a reporter molecule. Some proteins visualised by this method include neuroligin-1 in rat neuronal cultures (Howarth et al., 2006), and α -amino-3-hydroxy-5-methyl-4-isoxazolepropionate (AMPA) receptors in hippocampal neurons (Howarth et al., 2008; Howarth et al., 2005).

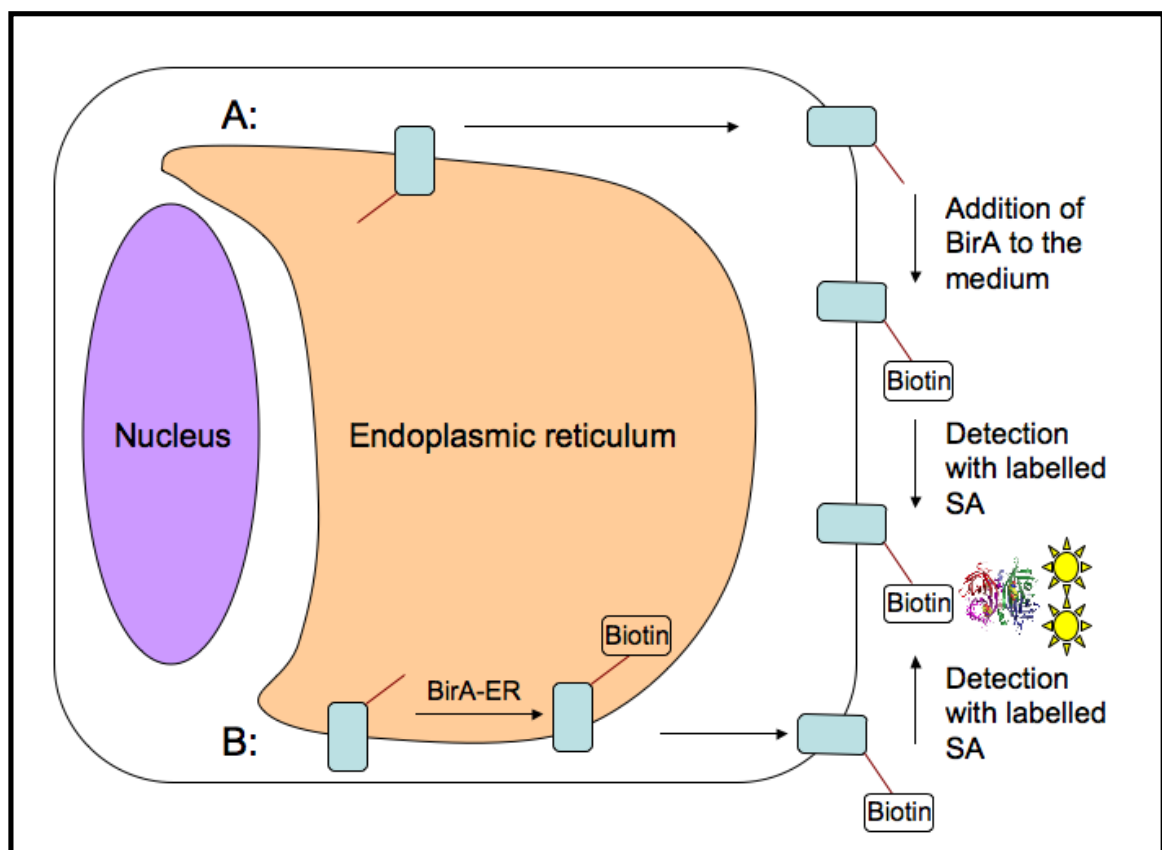


Figure 16. Cellular imaging with the BirA / AP system. There are two possible methods: method A which utilises recombinantly expressed BirA; and method B which utilises co-expressed BirA. The protein of interest (cyan rectangle) is expressed with the 15 aa AP tag (brown line). In method A, the tagged protein then traffics to the cell surface where BirA biotinylates the AP tag, enabling detection by SA labelled with a reporter molecule (yellow suns). In method B, the co-expressed BirA folds in the endoplasmic reticulum, where BirA biotinylates the AP tag in the protein of interest, which then traffics to the cell surface for detection with labelled SA. Diagram adapted from Howarth and Ting, 2008.

1.4.3 Detection and diagnostics

The SA-biotin interaction is one of the most widely used affinity-pairs in diagnostic applications, for isolation and identification of biotinylated molecules of interest. When used in affinity purification protocols, the extremely high affinity of the SA-biotin interaction enables highly stringent washing conditions to be used, leading to efficient capture of biotinylated molecules with high purity (de Boer et al., 2003). Also, there are few naturally biotinylated proteins in the cell, reducing the likelihood of false positives. The major drawback is that the SA-biotin interaction is not easily broken, and so elution of bound biotinylated molecules from a SA affinity column is extremely difficult. Harsh elution conditions can only be used if either the molecule of interest is resilient to the elution conditions used e.g. heating at 70 °C in non-ionic aqueous solutions can be used to release biotinylated DNA molecules from SA-coated beads (Holmberg et al., 2005), or if the molecule of interest is not needed in a functional form, e.g. boiling in SDS-containing buffer can be used if the downstream analysis of the isolated molecule involves, for example, mass spectrometry, and not a functional assay (Mach-Aigner et al., 2010). Otherwise, the biotin analogue 2-iminobiotin can be used (see Section 1.1.4), either as a tag to give pH-dependent capture and elution from a SA / Av affinity column (Orr, 1981), or as the capturing matrix in an affinity column for the isolation of SA-tagged molecules. Alternatively, SA can be used to capture proteins that have been genetically fused to linear peptides that bind to SA with micromolar to nanomolar affinities (see Section 1.1.5). Strep-tag (Schmidt and Skerra, 1993), Strep-tag II (Schmidt et al., 1996), Nano-tag₉ or

Nano-tag₁₅ (Lamla and Erdmann, 2004) all allow capture on a SA affinity column, with elution by competition with biotin or desthiobiotin.

In addition to the isolation of biotinylated molecules, SA is a common component of biosensors and protocols used for detection and identification of molecules. SA-coated biosensor chips and array slides are commercially available for the capture and immobilisation of biotin-labelled DNA, RNA and proteins in gene expression and protein profiling applications. SA-coated microplates for enzyme-linked immunosorbent assays (ELISAs) are also available, for capture of biotinylated molecules, such as biotinylated antigens or antibodies (SA conjugated to a reporter molecule also features in the final detection step of ELISAs for binding to the biotinylated detection antibody).

Other, more novel uses of SA in research have emerged in recent years. SA-coated microbeads have been used to achieve direct fluorescence detection of point mutations in human genomic DNA (Meng et al., 2010), as well as being used in conjunction with DNA bio-barcode for PCR-free quantitative detection of target nucleic acids (Zhu et al., 2008). The SA-biotin interaction has also been used to tether single-stranded DNA within an α -hemolysin protein pore in nanopore sequencing experiments (Stoddart et al., 2009).

1.4.4 Nanofabrication

Due to its highly stable biotin and biotin-conjugate binding, low non-specific binding and its multivalency, SA is also used as a molecular adaptor and building block in supramolecular assembly.

The SA-biotin interaction has been used as a building block to form many diverse, 'bottom-up' self-assembling nanostructures. Examples include periodic SA protein arrays using biotinylated oligonucleotides, which can be used to direct the assembly of other biotinylated molecules (Voigt et al., 2010; Yan et al., 2003) and monolayers of end-to-end-linked gold nanorods that can be used to direct assembly of higher-order arrays of nanomaterials (Caswell et al., 2003).

The SA-biotin interaction has also been used as a molecular adaptor to stably join one molecule to another. For example, the binding-pair have been used to couple protein cages to monoclonal antibodies (Suci et al., 2009), which could be utilised to deliver diagnostic or therapeutic agents to specific proteins or cell types. Another interesting use of SA is to increase the processivity of a DNA polymerase for use in 'sequencing by synthesis' (see also Chapter 4 discussion) (Williams et al., 2008). In this application, the strong biotin-binding ability and tetravalency of SA are utilised to form a 'clamp' across the binding cleft in the DNA polymerase, trapping the DNA in the polymerase and also enabling stable and controlled immobilization of the complex onto a surface.

1.5 Alternative tagging strategies for bioorthogonal chemistry

Tagging and detecting biomolecules using bioorthogonal chemistry has become an extremely useful tool for studying molecules in their native setting. The key requirements of applying bioorthogonal chemistry to living systems are that the components must react rapidly, be highly selective and not interfere with native biological processes (Sletten and Bertozzi, 2009). A typical tagging experiment utilising bioorthogonal chemistry involves using the cell's own synthetic apparatus to incorporate a unique chemical motif into the molecule of interest, which is then covalently modified with an exogenously delivered probe (Prescher and Bertozzi, 2005).

Because of their non-covalent binding, the SA-biotin pair is referred to as an interaction, rather than a bioconjugation reaction. However, the extremely stable biotin-binding has resulted in the SA-biotin interaction being considered a valuable tool in the vast bioorthogonal chemistry toolbox, being used in a variety of applications for tagging and stably joining objects together (see Section 1.4). Although purified proteins can be easily functionalised with biotin (by using a biotin derivative that has the valeryl chain functionalised with a reactive group to give covalent linkage to various functional groups), site-specific biotinylation of a protein of interest in its native surroundings makes use of the AP tag, a short chemical handle that can be genetically fused to the protein of interest and biotinylated *in vivo* by the exogenously-delivered biotin ligase enzyme BirA (see Section 1.4.2.1) (Howarth et al., 2005). There is a huge array of bioorthogonal chemical reactions and reporters being used to

study living systems, and excellent reviews of this field have already been published (Prescher and Bertozzi, 2005; Sletten and Bertozzi, 2009). It is beyond the scope of this thesis to comprehensively review this field here; instead, some of the more important tagging methods are discussed below, in comparison to the SA-biotin interaction.

1.5.1 Click chemistry

The term 'click chemistry' refers to a set of reactions that are simple to use, rapid, specific and give high yields (Kolb et al., 2001). The copper-catalysed 1,3-dipolar cyclo-addition reaction between an azide and alkyne (also known as the Huisgen cycloaddition) is a model 'click chemistry' reaction (Huisgen, 1963) (Figure 17). The Huisgen cycloaddition is widely used in many different applications, due to its versatility; the reaction can take place over wide temperature (0 to 160 °C) and pH (5 to 12) ranges (Hein et al., 2008), and with the addition of a copper catalyst, proceeds at a rapid rate (Rostovtsev et al., 2002). Also, the azide and alkyne functional groups can be easily incorporated into molecules and are highly stable (Bock et al., 2006). 'Click chemistry' reactions have been used to give site-specific labelling and conjugation of proteins; however, the reaction rate, even in the presence of a copper catalyst, is orders of magnitude slower than the diffusion limit and considerably slower than the SA-biotin on-rate, resulting in long incubation times being required when the reactants are present at low concentration (Link and Tirrell, 2003). The presence of copper can also be toxic to biological systems (Wolbers et al., 2006).

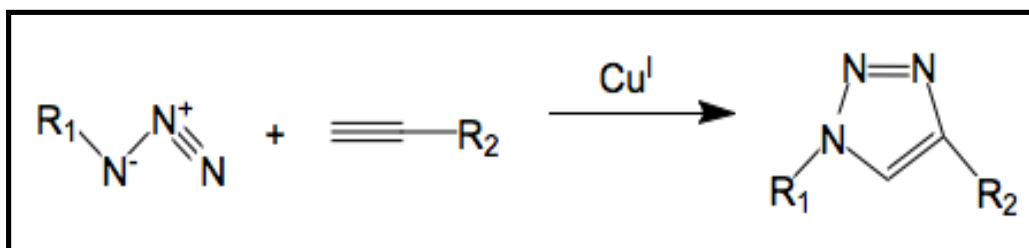


Figure 17. An example of a ‘click chemistry’ reaction: the copper-catalysed Huisgen 1,3-dipolar cycloaddition.

1.5.2 Native chemical ligation

Native chemical ligation (NCL) is another important application of chemistry to protein engineering, which results in the ligation of two peptide sequences by a peptide bond (Dawson et al., 1994). A peptide with a thioester at its C terminal undergoes nucleophilic attack by a N-terminal cysteine residue in another peptide, to give a thioester ligation product which spontaneously rearranges to give a peptide bond at the ligation site (Figure 18). This bioconjugation reaction is simple, rapid, proceeds in aqueous solution with a high yield and results in a stable peptide bond linkage between the two peptides. The main use of NCL is in the modification and synthesis of proteins (Sletten and Bertozzi, 2009), but it has also been used as a bioconjugation method, e.g. to immobilise polyethylene glycol on titanium surfaces (Byun et al., 2011). The SA-biotin interaction is not as stable as the amide linkage that results from NCL, but is also not dependent on a reducing environment for interaction, a limitation of NCL. Competition from endogenous cysteines will also impact on the success of NCL, which is limited to ligating only peptides together (Chen and Ting, 2005).

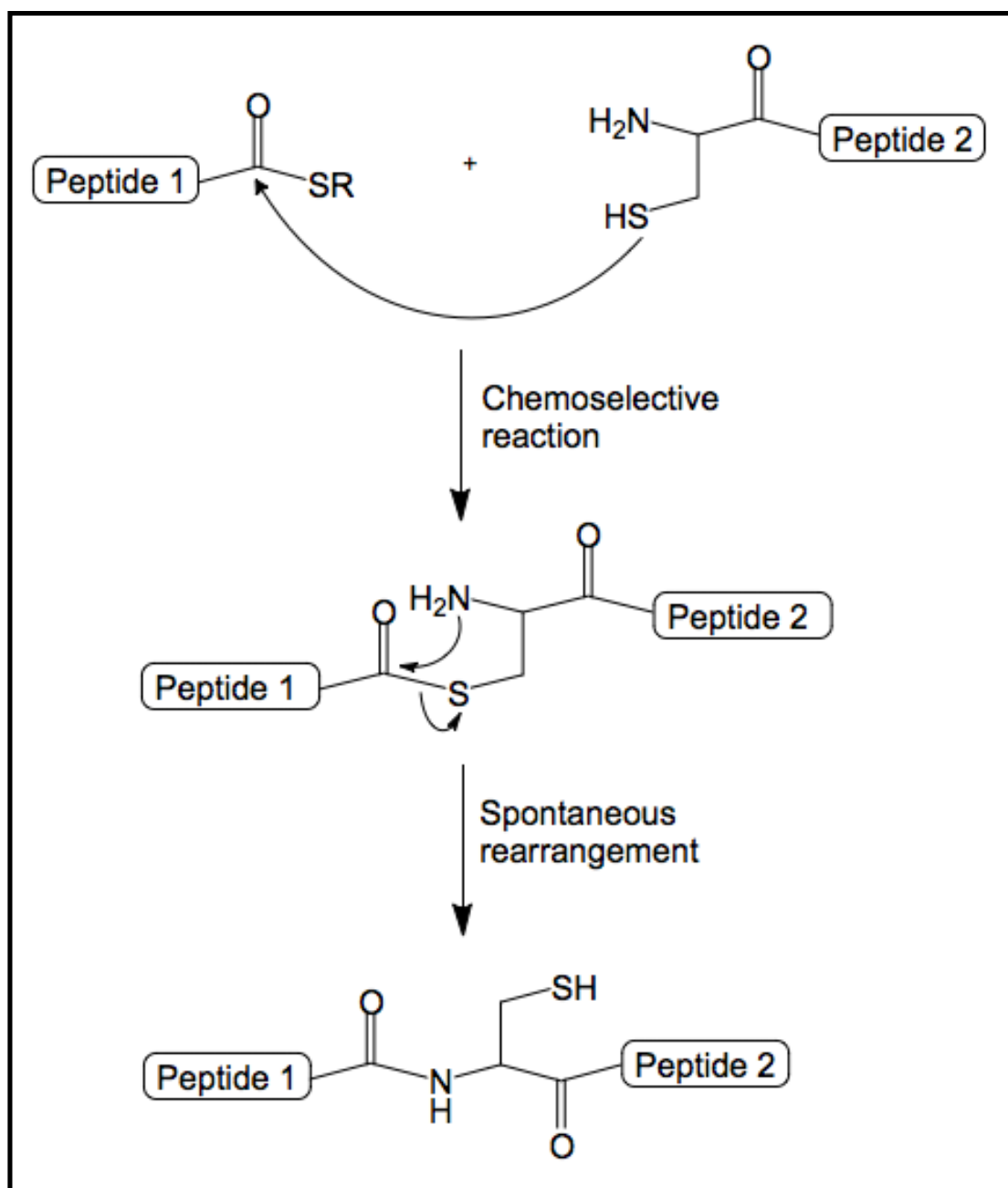


Figure 18. Native chemical ligation of two peptides. Peptide 1 contains a thioester at its C-terminal, which undergoes nucleophilic attack from the N-terminal cysteine of peptide 2, forming a thioester ligation product that spontaneously rearranges to give a native peptide bond between the two peptides.

1.5.3 Antibody-antigen binding

The affinities of antibodies generated *in vivo* tend to fall within the range 10^{-5} to 10^{-10} M (Kobayashi and Oyama, 2011). This is due to the existence of 'affinity ceilings', imposed by biological constraints, and a lack of selection pressure for antibodies that have much slower dissociation rates than the endocytosis rate into B cells (Foote and Eisen, 2000). However, *in vitro*, higher antibody affinities can be generated, as these limitations do not apply. For example, *in vitro* maturation methods have resulted in the production of antibodies with a dissociation constant of 5×10^{-14} M (Boder et al., 2000). However, just as an affinity ceiling exists *in vivo*, a methodological ceiling will exist *in vitro*, imposed by the inability to select for extremely slow dissociation rates (Foote and Eisen, 2000).

Antibody conjugates have found use in many different applications, including protein purification (e.g. anti-FLAG tag antibodies (Brizzard et al., 1994)), protein detection (e.g. in ELISAs and Western blotting (Renart et al., 1979)) and imaging biomolecules in biological systems (Massoud and Gambhir, 2003). Antibodies have also been the focus of a large number of engineering efforts to increase their utility (Hudson and Souriau, 2003). However, the size and charge of an antibody conjugate frequently prevents entry into the cell (Prescher and Bertozzi, 2005), and in general, the stability of antibody-antigen binding is much lower than SA-biotin binding.

Antibodies with 'infinite affinity' exist, where antibody-antigen binding is followed by a covalent reaction (Chmura et al., 2001). These 'infinite affinity'

antibodies are created by rational design, involving inspection of the structure of the antibody-antigen complex followed by the introduction of a nucleophile and partner electrophile at positions in the antibody and antigen where they will be brought into close proximity on binding, greatly increasing the local concentration and triggering a covalent reaction. Provided the rate of covalent bond formation is faster than the dissociation rate, an 'infinite affinity' antibody-antigen interaction is created. Whilst separate in solution, the low reactivity of these groups in the antibody and antigen give low cross-reactivity. In comparison to the natural antibody-antigen interaction (without the reactive groups), these 'infinite affinity' antibodies have an unchanged on-rate but a greatly decreased off-rate; indeed, a covalent bond between the antibody and its antigen can prevent any dissociation, giving an infinitely-bound complex (Butlin and Meares, 2006). This is advantageous over the SA-biotin interaction, which does suffer from, albeit very low, biotin dissociation. Also, antibodies can be humanised (Almagro and Fransson, 2008) to avoid triggering an immune response, which can limit the repeated use of SA (a bacterial protein) in living organisms (Butlin and Meares, 2006). However, this bioconjugation technique would be extremely time-consuming and labour-intensive to use as a general tagging system, as it involves inspection and modification of every antigen's antibody. Also, 'infinite affinity' antibodies often utilise a free thiol group to form the covalent linkage, which can be cross-reactive and easily oxidised.

1.5.4 HaloTag

The HaloTag is a modified bacterial haloalkane dehalogenase (34 kDa) which forms a covalent bond to its chloroalkane ligand (Los et al., 2008). The HaloTag can be genetically fused to the protein of interest and the chloroalkane ligand can be appended to whichever molecule is most suited to the application it is being used for, e.g. a fluorescent dye for protein imaging or a solid support for purification (Ohana et al., 2011). Advantages of using the HaloTag are that it is small, specific, does not need any cofactors for activity, and has an on-rate comparable to the SA-biotin interaction. However, the HaloTag protein is not as stable as SA, with optimum HaloTag binding being restricted to a narrow range of conditions, i.e. pH range 6 – 8.5. Also, the HaloTag is restricted to fusion at the N- or C-terminii of the protein of interest (Promega, 2009).

1.5.5 Dihydrofolate reductase

Dihydrofolate reductase (DHFR) is an essential enzyme that catalyses the NADPH-dependent reduction of dihydrofolate to tetrahydrofolate, which is required for the synthesis of purines and thymidylate. When inhibited, the cell cycle is halted leading to cell death, and so to exploit this, many DHFR inhibitors have been developed, for use as anticancer, antibacterial, antifungal and antiparasitic agents (Chan and Anderson, 2006). One particular inhibitor, methotrexate, binds to *E. coli* DHFR with a K_d of 25 pM (Sasso et al., 1994), an impressive affinity, but still weaker than the SA-biotin interaction. The DHFR-methotrexate interaction has been used in cellular imaging applications, where DHFR (157 aa) is genetically fused to the protein of interest and fluorescently-labeled methotrexate is used for visualisation (Miller et al., 2004). However, dissociation will occur, as with the SA-biotin interaction, due to the non-covalent nature of the interaction. Also endogenous DHFR in cells can give high background levels (Chen and Ting, 2005).

1.5.6 FIAsH

A small (663 Da) non-fluorescent arsenoxide molecule (Figure 19) developed by Tsien and coworkers (Griffin et al., 1998) can also be used in cellular imaging applications, as well as for labelling proteins *in vitro*. It is appropriately named FIAsH (fluorescein arsenical helix binder), as it becomes highly fluorescent on binding the tetracysteine motif CCXXCC (where XX can be any amino acids but is optimally proline and glycine (Adams et al., 2002b) (Figure 20)). Optimisation of the tetracysteine motif and FIAsH structure have produced a range of FIAsH derivatives capable of binding many different tetracysteine motifs with K_d values in the pM range (and can be as strong as 2 pM) (Adams et al., 2002b). The small size of the FIAsH reagent and its high-affinity binding to the tetracysteine motif make it a promising tool in protein labelling. However, the disadvantages of using FIAsH derivatives include: having high levels of non-specific binding, due to binding endogenous, cysteine-rich proteins; needing a reducing environment to label the tetracysteine motif; and possible toxicity due to the arsenic content (Griffin et al., 2000).

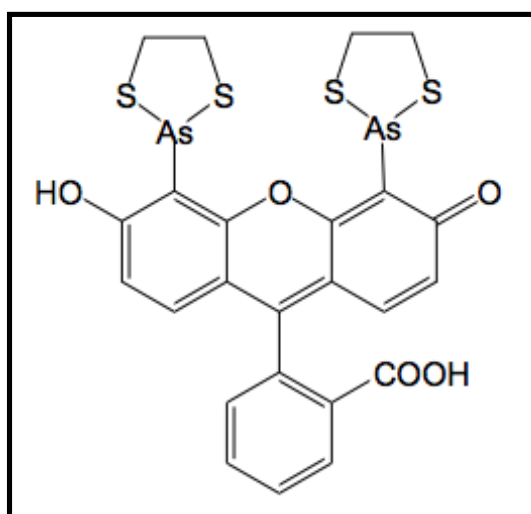


Figure 19. The FIAsH reagent consists of fluorescein with As(III) substituents at the 4'- and 5'- positions (Griffin et al., 1998).

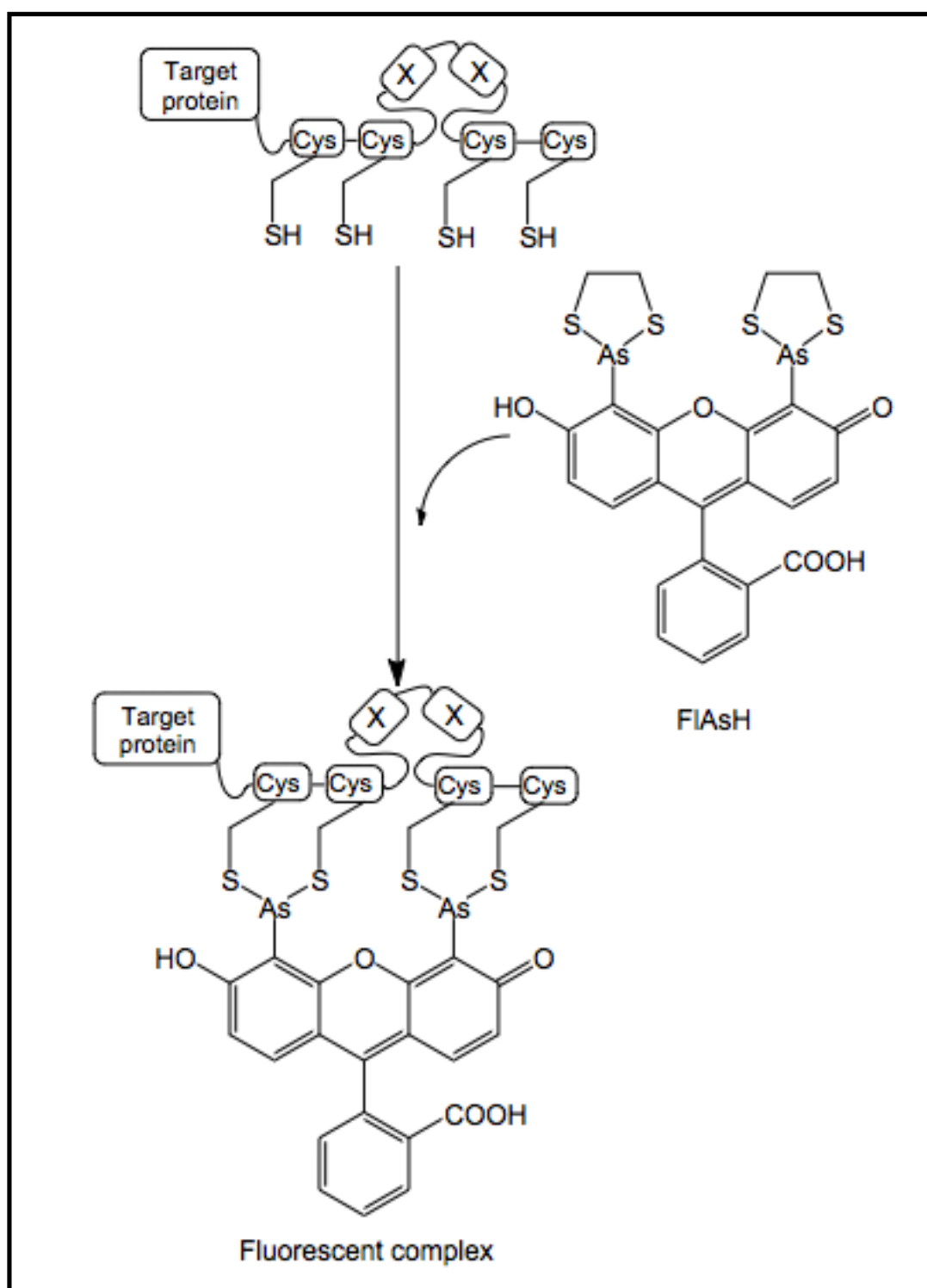


Figure 20. The binding of the FIAsh reagent to a tetracysteine motif in the target protein produces a highly fluorescent complex. The tetracysteine motif is shown as CCXXCC above, but is optimally CCPGCC in an α -helical conformation (Griffin et al., 2000).

1.6. Aims of this thesis

This thesis details efforts made at increasing the utility of the SA-biotin interaction. Section 1.3 highlights only a selection of the numerous engineering efforts that have been focussed on SA over the decades. However, to my knowledge, no group has produced and convincingly characterised a SA mutant with increased binding stability to *both* biotin and biotin-conjugates. There is a need for a higher-stability biotin-binder, and this is discussed in detail in Chapter 3. The results sections of this thesis (Chapters 3, 4 and 5) therefore discuss the production and characterisation of SA mutants with increased biotin-binding stability.

Rational site-directed mutagenesis of SA produced mutant proteins exhibiting a wide range of affinities towards biotin and biotin-conjugates (details given in Appendix A1), of which a subset possessed decreased dissociation rates from a biotin-conjugate. Chapter 3 details the production and characterisation of these higher-stability mutant SAs, as well as introducing the most promising mutant, which was named traptavidin (Tr). The further characterisation of Tr and its use in probing the force generated during translocation of the molecular motor FtsK are discussed in Chapter 4. In order to gain a deeper understanding of how the mutations in Tr exert their effect, the structures of both apo- and biotin-bound Tr were solved using x-ray crystallography and are presented in Chapter 5.

The results of this work are summarised in Chapter 6, where the future prospects of this research are also discussed. It is hoped that this thesis will

make a valuable contribution to the large body of research that already exists on the SA-biotin interaction.

The utility of SA can be increased in other ways in addition to increasing the biotin-binding stability, such as altering the valency or topology of SA. I have conducted provisional work into creating both monomeric SA and single-chain SAs (not included in this thesis) and the potential for these approaches to increase the functionality of SA is also discussed in Chapter 6. Increasing the ease and yield of recombinant SA expression should also increase the utility of SA, and provisional work towards this is detailed in Appendix A2.

Chapter 2: General Materials and Methods

This chapter gives an outline of the materials and methods used in this study.

More detailed protocols are given in specific chapters.

2.1 Sources of reagents and enzymes

Restriction endonucleases were purchased from New England Biolabs (UK). All DNA oligonucleotides were purchased from Invitrogen (UK) and their sequences are detailed in the 'materials and methods' section of the relevant chapter. All chemicals were purchased from Fisher Scientific (UK) unless otherwise stated.

2.2 Culture media

Luria Broth (LB) and 2x TY media were used for bacterial cell culture. They were prepared by the in-house media kitchen according to the recipes in Table 3 and sterilised by autoclaving for 20 minutes at 121 °C. Detailed culturing protocols and conditions used are detailed in the relevant chapters.

	LB media	2x TY media
Tryptone (g)	10	16
Yeast extract (g)	5	10
NaCl (g)	10	5

Dissolved in 1 L of sterile double-distilled H₂O and pH adjusted to pH 7.0

Table 3. Recipes for Luria Broth and 2x TY media for bacterial cell culture.

2.3 Competent bacterial strains and transformation

E. coli XL1-Blue supercompetent cells (Agilent Technologies, UK) were used in this study for the cloning of recombinant plasmids (Table 4). Three *E. coli* expression strains were used in this study: BL21 (DE3) RIPL cells (Agilent Technologies, UK); B834 (DE3) pLysS cells (EMD Chemicals, UK); and BL21 (DE3) pLysS cells (Agilent Technologies, UK). For recombinant strain selection, ampicillin (at 100 µg/mL) and chloramphenicol (at 25 µg/mL) were used.

Plasmids were subcloned into these strains using a heat-shock method. Competent cells were thawed on ice and 20 µL aliquots added to pre-chilled sterile 1.5 mL micro-centrifuge tubes. 50 ng of purified DNA plasmid or 4 µL of DpnI-digested PCR mix after site-directed mutagenesis was added to the cells, tubes flicked gently to mix and then incubated on ice for 20 minutes. The cells were then heat-shocked at 42 °C for 45 seconds in a water-bath followed by incubation on ice for 2 minutes. 300 µL of LB media (with no antibiotic selection) was then added to each tube and incubated at 37 °C for 1 hour. The entire transformation mixture was then plated on LB plates containing 1.5 % agar and 100 µg/mL ampicillin and incubated overnight at 37 °C. Freshly picked colonies were then used for small-scale culturing for DNA purification (from XL1-Blue cells) or large-scale culturing for protein purification (BL21 (DE3) RIPL, BL21 (DE3) pLysS or B834 (DE3) pLysS cells).

Strain	Genotype	Function
XL1-Blue	<i>recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F' proAB lacI^q ΔM15 Tn10 (Tet^R)]</i>	Plasmid replication
BL21 (DE3) RIPL	<i>B F⁻ ompT hsdS(r_B⁻ m_B⁻) dcm⁺ Tet^R gal λ(DE3) endA Hte [argU proL Cam^R] [argU ileY leuW Strep/Spec^R]</i>	Expression
B834 (DE3) pLysS	<i>F⁻ ompT hsdS(r_B⁻ m_B⁻) gal dcm met (DE3) pLysS (Cam^R)</i>	Expression
BL21 (DE3) pLysS	<i>B F⁻ ompT hsdS(r_B⁻ m_B⁻) dcm⁺ Tet^R gal λ(DE3) [pLysS Cam^R]</i>	Expression

Table 4. Genotype of competent bacterial strains used.

2.4 Site-directed mutagenesis

Detailed protocols, including PCR recipes and cycling programs, are given in the relevant chapters. DNA concentration was measured from a 2 μL sample using a NanoDrop 2000 (Thermo Scientific) spectrophotometer at 260 nm.

2.5 Agarose gel electrophoresis

DNA fragments were resolved using agarose gel electrophoresis through a 0.7 % (w/v) agarose gel containing 0.01 % (v/v) ethidium bromide in tris-acetate (TAE) buffer (40 mM Tris, 0.1 % (v/v) acetic acid and 1 mM EDTA). Samples were mixed with 6X DNA loading dye (1.2 M sucrose, 0.7 mM bromophenol blue, 0.9 M xylene cyanol FF) prior to loading and run alongside 2-log DNA ladder (New England Biolabs, UK) for size estimation. Gels were run submerged in TAE buffer at 130 V and DNA visualised using UV illumination at 254 nm.

2.6 Protein expression and purification

Detailed protein expression and purification protocols are given in the relevant chapters. Protein concentration was measured using a NanoDrop 2000 (Thermo Scientific) spectrophotometer at 280 nm.

2.7 Electrophoretic separation of proteins

Proteins were separated and visualised by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) (Laemmli, 1970). Gels consisted of a 5 % (v/v) acrylamide stacking layer (0.125 M Tris-HCl pH 6.8, 0.1 % (w/v) SDS and 5 % (v/v) acrylamide) and a variable separating layer, depending on separation required (0.38 M Tris-HCl pH 8.8, 0.1 % (w/v) SDS and 6 %, 8 % or 18 % (v/v) acrylamide). A 30 % (v/v) acrylamide : *N,N'*-methylenebisacrylamide mixture (37.5 : 1 ratio, Sigma, UK) was used to make the gels, which were poured in the lab using disposable plastic gel cassettes (Invitrogen, UK).

SDS-PAGE was carried out in a XCell *SureLock*® Mini-Cell apparatus (Invitrogen, UK) at 200 V using tris-glycine running buffer (25 mM Tris-HCl, 200 mM glycine, 0.1 % (w/v) SDS). The apparatus was carefully packed in ice to avoid excessive heating. Samples were mixed with 6X SDS loading dye (0.24 M Tris-HCl pH 6.8, 24 % v/v glycerol, 120 µM bromophenol blue, 0.23 M SDS) prior to gel loading and were loaded adjacent to PageRuler™ prestained protein ladder (Fermentas, UK) to estimate the molecular weight of protein bands. After running, resolved proteins were visualised by staining for

1 hour either with Coomassie Brilliant Blue solution (1.25 g Coomassie Brilliant Blue R250 in 500 mL of methanol : water : glacial acetic acid (4.5 : 4.5 : 1 v/v/v)), followed by overnight destaining in 30 % methanol (v/v), 10 % glacial acetic acid (v/v), or Instant Blue solution (Triple Red, UK).

2.8 Figure presentation

All figures containing 3D molecular structures were produced using PyMOL molecular visualisation system (Schrodinger, 2010). Chemical structures were displayed using ChemDraw software.

Chapter 3: Production and Characterisation of Higher-Stability Streptavidin Mutants

3.1 Introduction

As discussed in detail in Chapter 1, there exists a tremendous body of research on SA engineering. Considering SA's biotin-binding affinity, engineering efforts have attempted to lower or abrogate biotin-binding, produce tuneable biotin-binding, and impart novel binding abilities. High-throughput screening of SA libraries, together with the more labour- and time-consuming methods of rational mutagenesis and directed evolution, have produced over 200 mutants of SA (Laitinen et al., 2006). However, to my knowledge, no researchers have reported setting out with the intention of increasing SA's biotin-binding stability, perhaps because of the widely held assumption that the SA-biotin interaction has already been optimised by nature, with any mutation made to the system expected to reduce its performance. Two research groups have appeared to 'stumble upon' a SA mutant with higher than expected biotin-binding stability during their attempts to alter SA antigenicity (Meyer et al., 2001) or affinity towards a particular biotin analogue (Levy and Ellington, 2008). In a search for a SA with lower antigenicity, a few mutants were produced with up to 50 % slower biotin dissociation rates, although these rates were only determined once and no further investigation into the higher-stability mutants was undertaken (Meyer et al., 2001). A directed evolution approach was also used to produce a SA with increased binding affinity for desthiobiotin, and the mutant SA produced retained a high biotin-binding stability, with only a ten-fold faster biotin

dissociation rate than wildtype SA (Levy and Ellington, 2008). An explanation for this was only suggested recently (Magalhaes et al., 2011), and is discussed further in Chapter 5.

3.1.1 The need for a higher-stability biotin-binder

All work presented in this body of research was directed towards increasing the utility of SA. This chapter presents work towards increasing the binding-stability of SA. Chapter 1 discussed the remarkable properties of the SA-biotin interaction, such as the rapid on-rate (Hyre et al., 2006), extremely slow off-rate (Levy and Ellington, 2008), extreme thermal stability and resistance to denaturants (Gonzalez et al., 1999). However, despite the apparent stability of biotin-binding, the SA-biotin interaction can and does fail under certain experimental conditions.

For example, the SA-biotin interaction is often used in cellular imaging applications (see Section 1.4.2), with a labelled SA detecting the biotinylated protein of interest. However, the lower pH of the endosomes, together with the higher temperature (37 °C) used for live cell imaging, can increase SA-biotin dissociation, limiting the time for which the protein can be detected. This was indeed the case when the SA-biotin interaction was used to visualise the cellular trafficking of nicotinic acetylcholine receptors; the receptors could only be visualised for ~ 2 hours before SA-biotin dissociation started to occur, whilst the lifetime of the receptors was ~ 4 days (Bruneau et al., 2005).

Nanoparticle attachment has also been shown to result in a substantial decrease in SA-biotin stability (Swift et al., 2006); the K_d for a biotinylated peptide increased approximately a million-fold when streptavidin was attached to beads (Buranda et al., 1999). This is extremely relevant as SA is frequently attached to nanoparticles in many different applications, such as: imaging, where SA can be attached to quantum dots; and purification protocols, where SA can be immobilised on sepharose beads in a column or attached to magnetic beads for facile isolation. Also, investigations using a laminar flow chamber revealed that in the presence of shear-forces lower than those in a blood capillary, SA-coated beads do not attach to a biotinylated surface but instead roll across, becoming bound and arresting for periods of 20 ms to tens of seconds (Pierres et al., 2002). This has implications for the use of SA in targeted radionuclide therapy, where radiolabelled SA is used to bind to biotinylated antibodies raised against a tumour cell antigen (Boerman et al., 2003).

Furthermore, in molecular motor investigations, SA has failed to prevent the translocation of motors such as helicases, RNA polymerase or DNA polymerase along DNA (Morris et al., 2001), suggesting the binding stability of the SA-biotin interaction is not sufficient to withstand the translocation force of these motors.

Also, the high thermal stability of SA's biotin-binding, together with the high thermal stability of the SA tetramer, has enabled the SA-biotin interaction to be used in applications occurring at high temperatures, such as in PCR (Williams et al., 2008), BEAMing (Dressman et al., 2003) and 454 DNA

sequencing (Margulies et al., 2005). However, the elevated temperatures in these applications do increase SA-biotin dissociation, resulting in the protocols having to be adapted to compensate for this increased dissociation, e.g. DNA has to be bis-biotinylated to reduce dissociation in BEAMing applications (Dressman et al., 2003) and the temperature of the elongation step has to be reduced to 54 °C (from the polymerase's optimum temperature of 74 °C) in PCR applications (Williams et al., 2008).

3.1.2 Aims of this chapter

This chapter details my work towards producing a SA mutant or mutants that can bind more stably to biotin and any biotin-conjugate. Rational site-directed mutagenesis was used; mutations in residues that are positioned close to the ureido or thiophene rings of biotin were avoided as these have been shown to invariably impede biotin binding (Klumb et al., 1998), and instead, mutations of residues adjacent to the biotin carboxyl and in the loop connecting β -strands 3 and 4 (L3/4) were explored. L3/4 was chosen as the focus for the mutagenesis as this flexible region is found disordered and in an 'open' conformation in apo-SA, but is stabilised and closes over the biotin-binding pocket in a 'closed' conformation on biotin-binding (Freitag et al., 1997).

Eight mutants were found to possess a slower dissociation rate from the fluorescent biotin-conjugate biotin-4-fluorescein (B4F). These eight mutants contained mutations at the Ser-52, Arg-53 and Ser-112 residues, in various combinations, resulting in the production of single, double and triple mutants. All eight mutants contained one or more mutations in L3/4; the possible mode

of action of these mutations is discussed here and in Chapter 5. Although rationally predicted to have increased biotin-binding stability, another 19 mutants had an increased B4F dissociation rate (and therefore decreased binding stability); the data on these weaker mutants are presented in Appendix A1.

The production and characterisation of the eight higher-stability mutants is presented in this chapter, including the B4F dissociation and association rates, thermostability and yields when expressed and purified in *E. coli*. These properties were taken into account when choosing the most promising mutant to further characterise (in Chapter 4).

3.2 Materials and methods

3.2.1 Site-directed mutagenesis

Rational site-directed mutagenesis was conducted on the residues in and around the biotin-binding site in SA, focussing on the residues in L3/4. The eight higher-stability SA mutants discussed in this chapter contained the mutations: S52G, S112D, S52G R53N, S52G R53S, S52G R53T, S52G S112D, S52G R53S S112D and S52G R53D.

Appropriate primers were designed to introduce the desired mutations into wildtype SA and their sequences are given in Table 5. Wildtype SA here refers to core SA with a His₆ tag at the C terminus, with the gene encoding the protein in the pET21a(+) plasmid. PCR was carried out in 25 μ L reactions in a Bio-Rad DNA Engine® Peltier Thermal Cycler. A typical reaction would contain the following:

Purified wildtype streptavidin template plasmid	25 ng
Forward primer	200 pmol
Reverse primer	200 pmol
10x Pfu reaction buffer	2.5 μ L
10 mM dNTPs	0.5 μ L
Pfu Turbo DNA polymerase	0.5 μ L
ddH ₂ O	to 25 μ L

Typical reaction conditions would be 95 °C for 2 minutes, followed by 20 cycles of 95 °C for 30 s, 55 °C for 1 minute and 68 °C for 6 minutes, followed by a final incubation at 68 °C for 5 minutes. 5 µL of the amplified reaction mixture underwent 0.7 % (w/v) agarose gel electrophoresis to visualise PCR products. 10 µL of the amplified reaction mix was digested with 10 units of DpnI (New England Biolabs, UK) at 37 °C for 1 hour to digest methylated and hemimethylated DNA. 4 µL of the digested mixture was then transformed into competent *E. coli* XL1-Blue cells (see Section 2.3) and transformants were selected on LB agar plates (1 % tryptone, 0.5 % yeast extract, 1 % NaCl, 1.6 % agar) supplemented with 100 µg/mL ampicillin (Amp). Individual colonies were picked and used to inoculate 12 mL of LB containing 100 µg/mL Amp, which was cultured overnight at 37 °C in a shaking incubator. Cells were then harvested by centrifugation at 4,000 g for 5 minutes and DNA extracted using a Qiagen Miniprep kit (Valencia, CA). Sequences were confirmed by sequencing with T7 reverse primer (Table 3) by Geneservice Ltd., UK.

3.2.2 Protein expression and purification

Plasmids containing desired mutations were subcloned into competent *E. coli* BL21 (DE3) RIPL cells (see Section 2.3) and transformants selected on LB-Amp plates. Individual freshly-grown colonies were picked and used to inoculate 12mL LB-Amp, incubated at 37 °C overnight. These starter cultures were then diluted 1:100 into large-scale LB-Amp cultures, which were grown to OD₆₀₀ 0.9 at 37 °C, induced with 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and incubated for a further 4 hr at 37 °C. Cells

were then harvested by centrifugation at 7,500 g for 15 min at 4 °C and pellets stored at -80 °C until induction was confirmed on 18 % SDS-PAGE.

Insoluble SA mutant proteins were purified according to a protocol adapted from that published previously (Howarth and Ting, 2008). Inclusion bodies were isolated from the cell pellet of a 750 mL culture by incubation with 10 mL 300 mM NaCl, 50 mM Tris, 5 mM EDTA, 0.8 mg/mL lysozyme, 1 % Triton X-100 pH 7.8 for 30 min at 25 °C on a platform rocker. The inclusion bodies were then sonicated for 9 min (pulsed) on ice at 40 % amplitude on a Sonics Vibra-Cell sonicator. Following centrifugation at 27,000 g for 15 min at 4 °C, the inclusion body pellet was washed three times in 10 mL 100 mM NaCl, 50 mM Tris, 0.5 % Triton X-100 pH 7.8 and then solubilised in 4 mL of 6 M guanidine hydrochloride (GuHCl) pH 1.55. Denatured protein in GuHCl was then centrifuged at 13,000 g for 15 min at 4 °C, and the clear supernatant refolded by rapid dilution into 170 mL of forcefully stirring PBS (137 mM NaCl, 2.7 mM KCl, 8.1 mM Na₂HPO₄ · 2H₂O, 1.8 mM KH₂PO₄ pH 7.4) at 4 °C, followed by stirring overnight. Refolded protein was then centrifuged at 17,700 g for 15 min at 4 °C to precipitate any misfolded proteins and 1 mL of packed Qiagen Ni-NTA Superflow resin, equilibrated in 300 mM NaCl, 50 mM Tris pH 7.8, was then added to the supernatant and rotated overnight at 4 °C. The next day, the resin was isolated by centrifugation, washed once with 5 mL 300 mM NaCl, 50 mM Tris, 30 mM imidazole pH 7.8 and then added to a poly-prep column (Bio-Rad, UK) for elution with 5 mL 300 mM NaCl, 50 mM Tris, 200 mM imidazole pH 7.8. The eluate was then dialysed three times against PBS and protein concentration and yield determined from the

absorbance at 280 nm (a 1 mg/mL solution of wildtype SA has a predicted A_{280} of 2.948 according to ProtParam (Gasteiger et al., 2005)). Protein purity was confirmed on SDS-PAGE. Working solutions were stored at 4 °C, with aliquots kept at -80 °C for long-term storage.

3.2.3 Biotin-4-fluorescein off-rate assay

The biotin-4-fluorescein (B4F) off-rate assay was conducted on both wildtype SA and every mutant made as a rapid and facile method of characterising the biotin-conjugate binding stability. The assay was conducted using a PHERAstar Plus plate-reader with 480 nm excitation and 520 nm emission (BMG LABTECH, UK). In this assay the binding of B4F (Figure 21) to an excess of SA results in quenching of the fluorescein emission (Kada et al., 1999). As the B4F dissociates, the fluorescence recovers, and this fluorescence recovery was measured every 10 minutes over 12 hours at 37 °C. The assay was performed using excess biotin, so that empty binding sites that form on B4F dissociation are immediately re-filled by biotin.

Wildtype or mutant SA, at a concentration of 1 μ M in 10 μ L PBS, was added to 12 nM B4F (Invitrogen, UK) with 0.12 mg/mL bovine serum albumin (BSA) in 170 μ L PBS and incubated for 1 hr at 37 °C. Fluorescence measurements were then immediately started after addition of either 20 μ L PBS or 20 μ L of 1 mM biotin in PBS. Measurements, in triplicate, were taken at 37 °C every 10 minutes for 12 hours. Percentage dissociation was calculated as (signal with biotin – signal without biotin)/(signal without quenching – signal without biotin) multiplied by 100. For the 'signal without quenching', no SA was added to the

B4F, enabling any bleaching of the B4F during the experiment to be observed. The 'signal without biotin' allowed the fluorescence readings to be corrected for the residual background fluorescence arising from any unbound B4F. Means and standard deviations were calculated from the triplicate measurements.

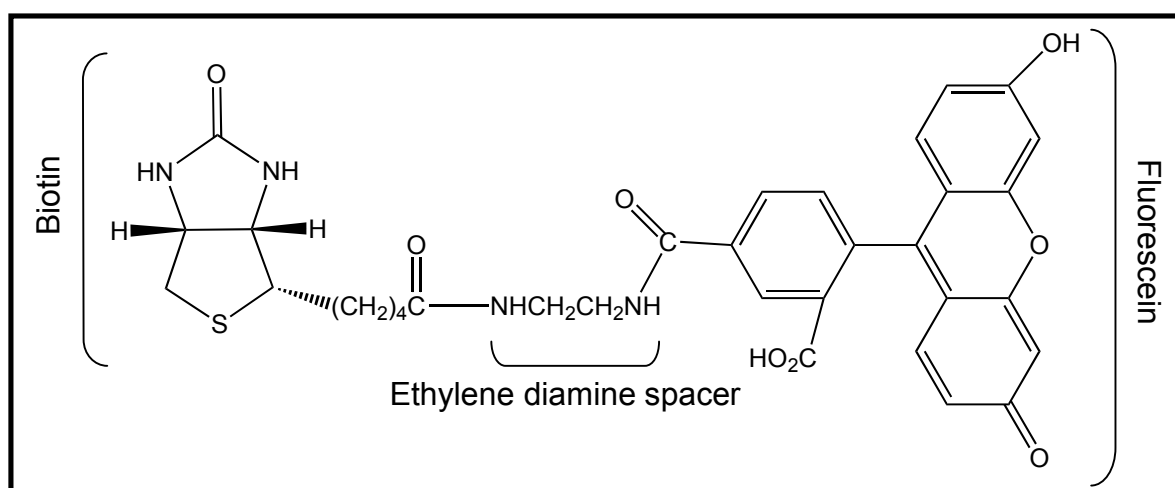


Figure 21. Biotin-4-fluorescein structure.

3.2.4 Biotin-4-fluorescein on-rate assay

In this assay the binding of B4F to a mutant protein or wildtype SA results in quenching of the fluorescein emission (Kada et al., 1999). This decrease in fluorescence was measured on a PHERAstar Plus plate-reader (BMG LABTECH, UK) with 480 nm excitation and 520 nm emission. The assay was conducted at 25 °C as the on-rate was too rapid to measure using this method at 37 °C. Fluorescence measurements were immediately started after 20 µL of 10 nM SA was added to 180 µL of 56 pM B4F, with triplicate readings taken every 6 s. The concentration of free B4F was calculated as (signal with B4F – signal without B4F)/(signal without protein – signal without B4F), multiplied by the starting [B4F], 50 pM. The 'signal without biotin' allowed the fluorescence readings to be corrected for the residual background

fluorescence arising from any unbound B4F. Linear regression using GraphPad Prism (GraphPad Software) was applied to the plot of $\ln$ [free B4F] against time, with the gradient equal to $k_{\text{on}} \times [\text{protein}]$. Error bars for the on-rate were calculated using the “LINEST” linear least squares curve-fitting routine in Microsoft Excel.

3.2.5 Thermostability assay

To determine tetramer thermostability, 3 μM of wildtype or mutant SA in PBS was heated at the indicated temperature for 3 min followed by cooling to 10 °C in a Bio-Rad DNA Engine® Peltier Thermal Cycler and then immediately placed on ice (Bayer et al., 1996). Samples were then run on 18% SDS-PAGE. The 100% monomer positive control was mixed with SDS loading buffer prior to heating at 95 °C for 3 min.

Name	Sequence (5' – 3')
S52G F S52G R	GTTGGTAACGCTGAAGGTAGATACGTTCTGACCGGTC GACCGGTCAGAACGTATCTACCTTCAGCGTTACCAAC
S52G R53N F S52G R53N R	GAATCCGCTGTTGGTAACGCTGAAGGCAACTACGTTCTGACCGGTCG CGACCGGTCAGAACGTAGTTGCCTTCAGCGTTACCAACAGCGGATTC
S52G R53S F S52G R53S R	GTTGGTAACGCTGAAGGTAGCTACGTTCTGACCGGTCG CGACCGGTCAGAACGTAGCTACCTTCAGCGTTACCAAC
S52G R53D F S52G R53D R	GAATCCGCTGTTGGTAACGCTGAAGGCGATTACGTTCTGACCGGTCGTTAC GTAACGACCGGTCAGAACGTAATCGCCTTCAGCGTTACCAACAGCGGATTC
S52G R53T F S52G R53T R	CCGCTGTTGGTAACGCTGAAGGCACCTACGTTCTGACCGGTCG CGACCGGTCAGAACGTAGGTGCCTTCAGCGTTACCAACAGCGG
S112D F S112D R	CCAGTGGTTGTTGACCGATGGCACCACCGAAGCTAAC GTTAGCTTCGGTGGTGCCATCGGTCAACAACCACTGG
T7R	GCT AGT TAT TGC TCA GCG G

Table 5. DNA oligonucleotides used in this study

3.3 Results

3.3.1 Protein production

Site-directed mutagenesis was used to produce the eight higher-stability SA mutants discussed in this chapter (S52G, S112D, S52G R53N, S52G R53S, S52G R53T, S52G S112D, S52G R53S S112D and S52G R53D), together with the 19 lower-stability mutants described in Appendix A1. The amino acid sequences of the higher-stability mutants are shown in Figure 22.

Streptavidin	AEAGITGTWYNQLGSTFIVTAGADGALTGTYESAVGNAESRYVLTGRYDSAPATDGSCTA
S52G	AEAGITGTWYNQLGSTFIVTAGADGALTGTYESAVGNAEGRYVLTGRYDSAPATDGSCTA
S112D	AEAGITGTWYNQLGSTFIVTAGADGALTGTYESAVGNAESRYVLTGRYDSAPATDGSCTA
S52G R53D	AEAGITGTWYNQLGSTFIVTAGADGALTGTYESAVGNAEGRYVLTGRYDSAPATDGSCTA
S52G R53N	AEAGITGTWYNQLGSTFIVTAGADGALTGTYESAVGNAEGRYVLTGRYDSAPATDGSCTA
S52G R53S	AEAGITGTWYNQLGSTFIVTAGADGALTGTYESAVGNAEGRYVLTGRYDSAPATDGSCTA
S52G R53T	AEAGITGTWYNQLGSTFIVTAGADGALTGTYESAVGNAEGRYVLTGRYDSAPATDGSCTA
S52G S112D	AEAGITGTWYNQLGSTFIVTAGADGALTGTYESAVGNAEGRYVLTGRYDSAPATDGSCTA
S52G R53S S112D	AEAGITGTWYNQLGSTFIVTAGADGALTGTYESAVGNAEGRYVLTGRYDSAPATDGSCTA

Streptavidin	LGWTVAWKNNYRNAHSATTWSGQYVGGAEARINTQWLLTSGTTEANAWKSTLVGHDTFTK
S52G	LGWTVAWKNNYRNAHSATTWSGQYVGGAEARINTQWLLTSGTTEANAWKSTLVGHDTFTK
S112D	LGWTVAWKNNYRNAHSATTWSGQYVGGAEARINTQWLLTSGTTEANAWKSTLVGHDTFTK
S52G R53D	LGWTVAWKNNYRNAHSATTWSGQYVGGAEARINTQWLLTSGTTEANAWKSTLVGHDTFTK
S52G R53N	LGWTVAWKNNYRNAHSATTWSGQYVGGAEARINTQWLLTSGTTEANAWKSTLVGHDTFTK
S52G R53S	LGWTVAWKNNYRNAHSATTWSGQYVGGAEARINTQWLLTSGTTEANAWKSTLVGHDTFTK
S52G R53T	LGWTVAWKNNYRNAHSATTWSGQYVGGAEARINTQWLLTSGTTEANAWKSTLVGHDTFTK
S52G S112D	LGWTVAWKNNYRNAHSATTWSGQYVGGAEARINTQWLLTSGTTEANAWKSTLVGHDTFTK
S52G R53S S112D	LGWTVAWKNNYRNAHSATTWSGQYVGGAEARINTQWLLTSGTTEANAWKSTLVGHDTFTK

Streptavidin	VKPSAAS
S52G	VKPSAAS
S112D	VKPSAAS
S52G R53D	VKPSAAS
S52G R53N	VKPSAAS
S52G R53S	VKPSAAS
S52G R53T	VKPSAAS
S52G S112D	VKPSAAS
S52G R53S S112D	VKPSAAS

Figure 22. Multiple sequence alignment of the eight higher-affinity mutants characterised in this chapter. The alignment was carried out using ClustalW. The conserved residues are highlighted by an asterisk and the eight β -strands indicated by the arrows are positioned according to the structure of SA. Mutated residues are coloured red.

3.3.2 Protein expression & purification

All eight SA mutants discussed in this chapter (S52G, S112D, S52G R53N, S52G R53S, S52G R53T, S52G S112D, S52G R53S S112D and S52G R53D), as well as wildtype SA, were expressed insolubly in *E. coli*, forming inclusion bodies. After conducting expression trials (data not shown), it was found that expressing in the BL21 (DE3) RIPL cell line with a four hour induction at 37 °C using 0.5 mM IPTG gave strong induction (Figure 23).

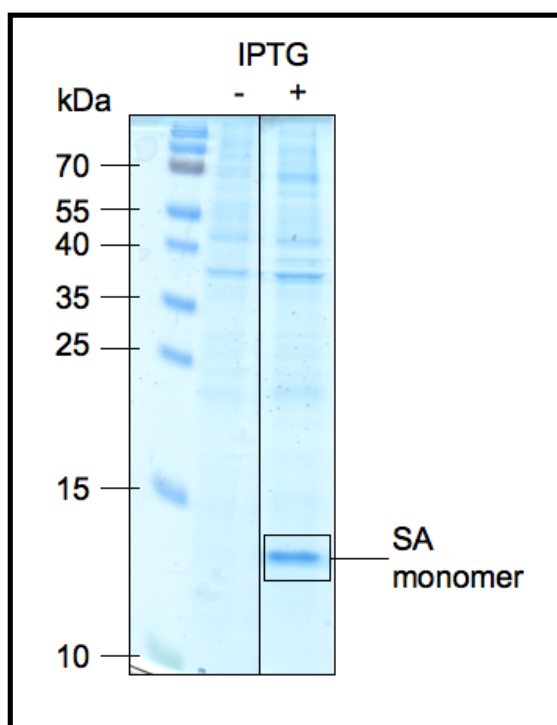


Figure 23. A four hour induction at 37 °C in BL21 (DE3) RIPL cells gave strong induction of SA. 10 µl samples of SA expression cultures before and after IPTG induction were boiled with 6X SDS loading buffer and run on 18 % SDS-PAGE.

As all proteins were expressed insolubly, they were isolated from inclusion bodies and then refolded *in vitro* by rapid dilution into PBS. On centrifugation after refolding, they gave different amounts of precipitate of misfolded protein, as some mutants were more amenable to *in vitro* refolding than others, and this is reflected in the yields obtained (Table 6).

Protein	Yield (mg of protein / litre of culture)
Wildtype streptavidin	10.3 ± 4.3
S52G	0.8 ± 0.2
S112D	0.3 ± 0.1
S52D R53N	5.6 ± 0.5
S52G R53S	2.1 ± 0.7
S52G R53T	0.5 ± 0.1
S52G S112D	1.4 ± 0.7
S52G R53S S112D	4.7 ± 1.2
S52G R53D	9.3 ± 2.7

Table 6. Protein yields after *in vitro* refolding by rapid dilution into PBS. Values shown are means of three refolding attempts ± 1 s.d.

On completing every purification, the proteins were run on SDS-PAGE, in order to evaluate the purity of the product. Boiled proteins were run on 18 % SDS-PAGE whilst unboiled proteins were run on 8 % SDS-PAGE. Example gels are shown in Figure 24.

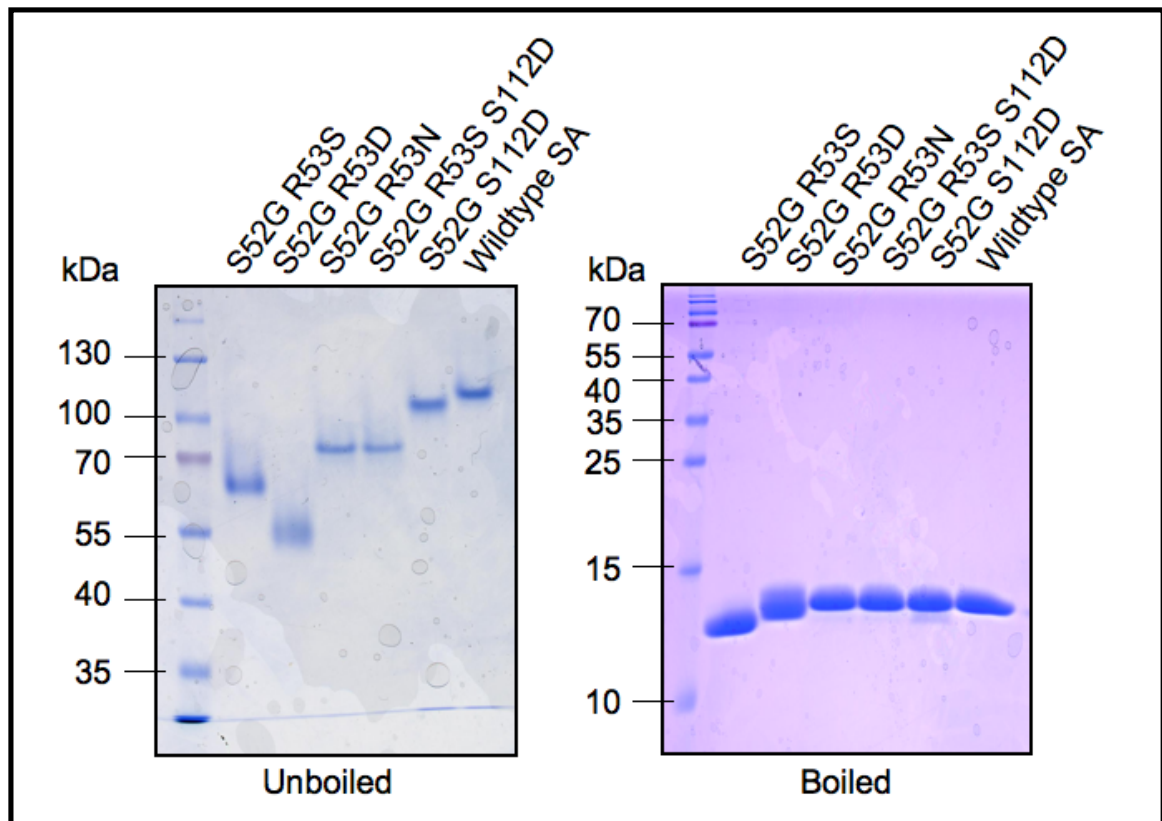


Figure 24. Protein purity was confirmed at the end of each purification by SDS-PAGE. Samples were mixed with 6X SDS loading buffer and run on 8 % SDS-PAGE to reveal the native tetrameric structure of the proteins (left). Samples were also boiled in 6X SDS loading buffer prior to 18 % SDS-PAGE to visualise the denatured monomeric proteins (right).

3.3.3 Biotin-4-fluorescein off-rates

The biotin-4-fluorescein (B4F) off-rate assay was conducted on every SA mutant made, in order to see how the mutant's B4F off-rate compares to that of wildtype SA. On binding to an excess of wildtype or mutant SA, the fluorescein emission of B4F is quenched. As the B4F dissociates, the fluorescence recovers, and this fluorescence recovery was measured over time at 37 °C, to visualise the B4F dissociation. This was a simple and rapid assay to evaluate a mutant's B4F off-rate compared to wildtype SA, and identified mutants with either a comparable or lower dissociation rate to SA for further characterisation, and, in some cases, further rounds of mutagenesis. In total, out of 27 mutants made, eight possessed lower B4F dissociation rates than wildtype streptavidin (Figure 25).

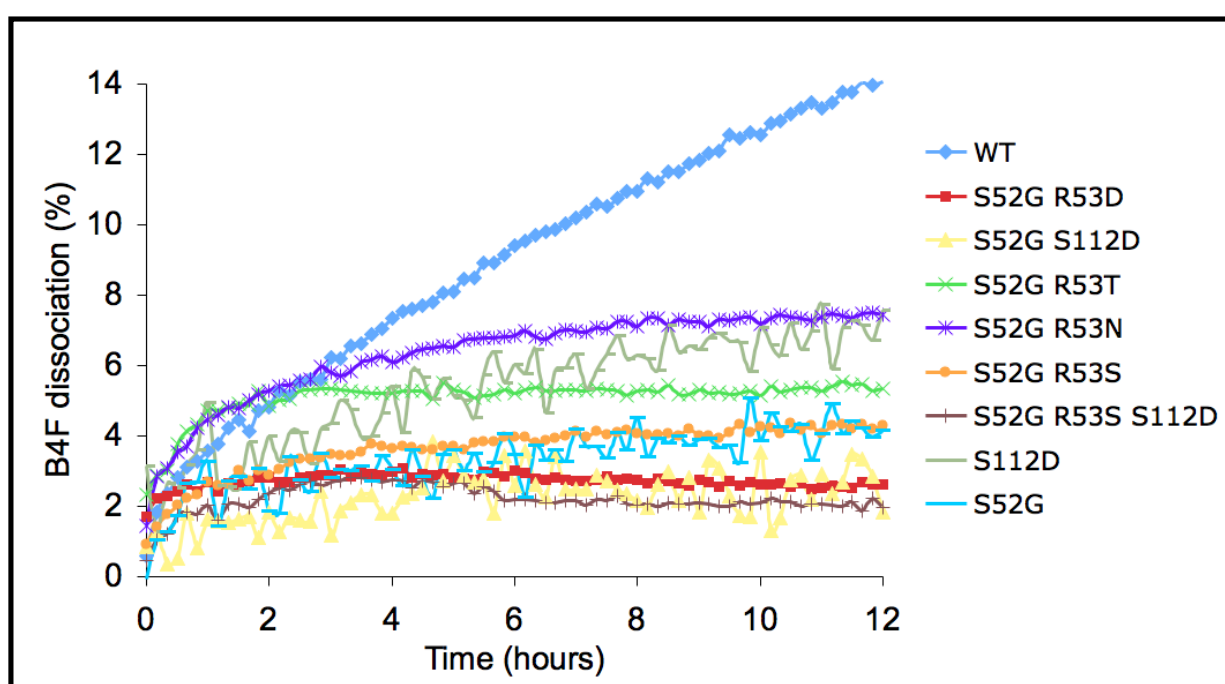


Figure 25. Eight streptavidin mutants had lower B4F off-rates than wildtype streptavidin. All off-rate assays were conducted at 37 °C in PBS with 0.12 mg/mL BSA. Data shown here without error bars for clarity, but error bars are shown for subsequent pair-wise comparisons.

The percentage quenching of B4F fluorescence obtained at the beginning of the assay before addition of excess free biotin is shown for each mutant in Table 7.

Protein	Quenching (%)
Wildtype SA	89.7 $\pm$ 0.6
S52G	71.0 $\pm$ 1.9
S112D	67.5 $\pm$ 0.2
S52D R53N	90.4 $\pm$ 0.3
S52G R53S	90.5 $\pm$ 0.4
S52G R53T	91.7 $\pm$ 0.1
S52G S112D	87.5 $\pm$ 1.5
S52G R53S S112D	90.9 $\pm$ 0.1
S52G R53D	90.0 $\pm$ 0.7

Table 7. Percentage quenching of B4F fluorescence by mutant SA binding. Means of triplicate measurements shown $\pm$ 1 s.d.

For ease of comparison, the B4F dissociation data for each mutant was plotted individually, alongside the data for wildtype SA (Figure 26).

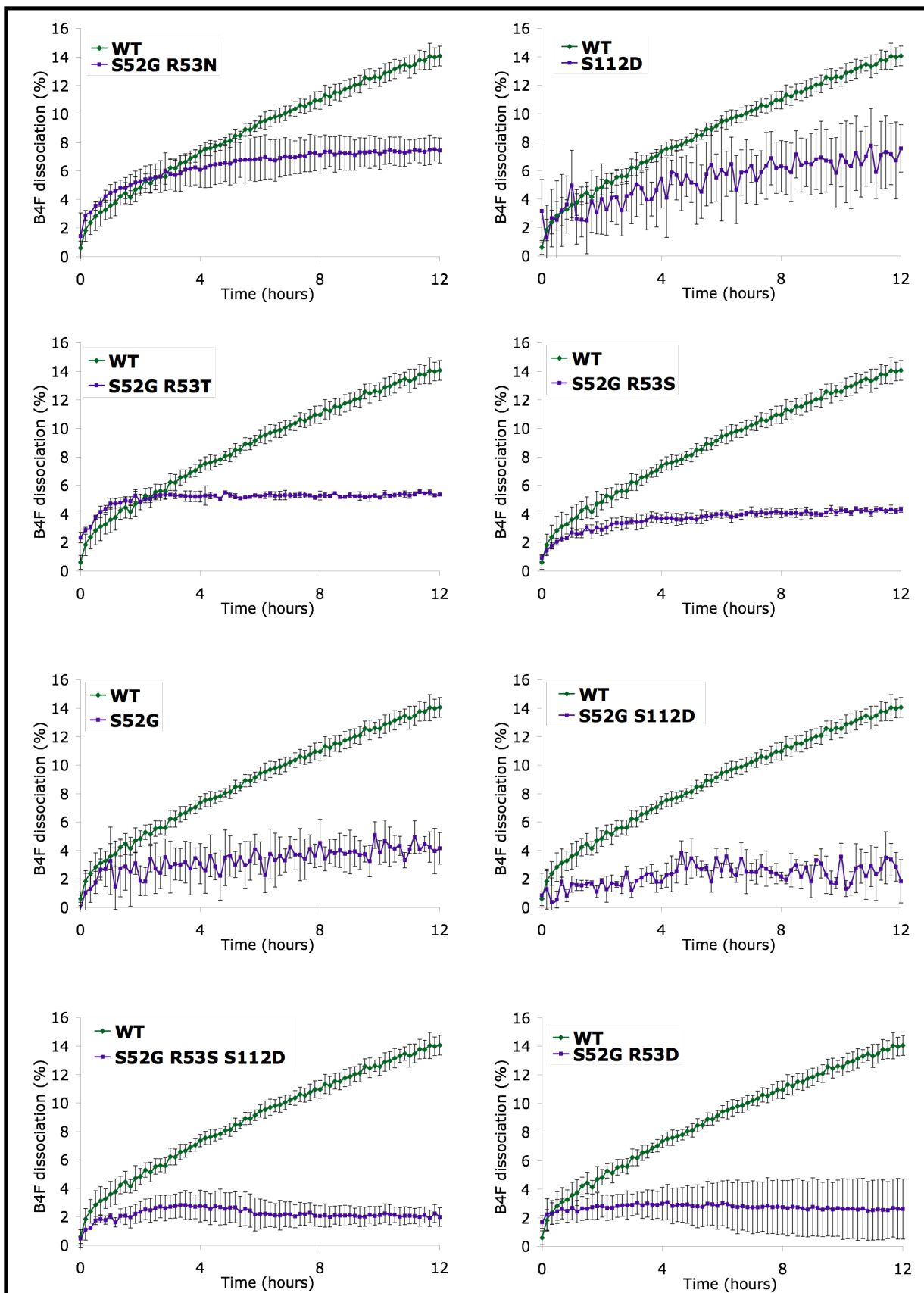


Figure 26. B4F off-rate data for each higher-stability mutant compared to wildtype streptavidin. Assays were conducted at 37 °C in PBS with 0.12 mg/mL BSA. Means of triplicate experiments shown, error bars $\pm$ 1 s.d. Some error bars are too small to be visible.

3.3.4 Biotin-4-fluorescein on-rates

The B4F off-rate assay was then altered to enable the on-rate to be measured (Figure 27). On binding to an excess of wildtype or mutant SA, the fluorescein emission of B4F is quenched and this decrease in fluorescence is measured over time to visualise the B4F binding. The on-rates were too rapid to measure at 37 °C (the temperature used for the B4F off-rate assays), and so the on-rate assay was conducted at 25 °C instead.

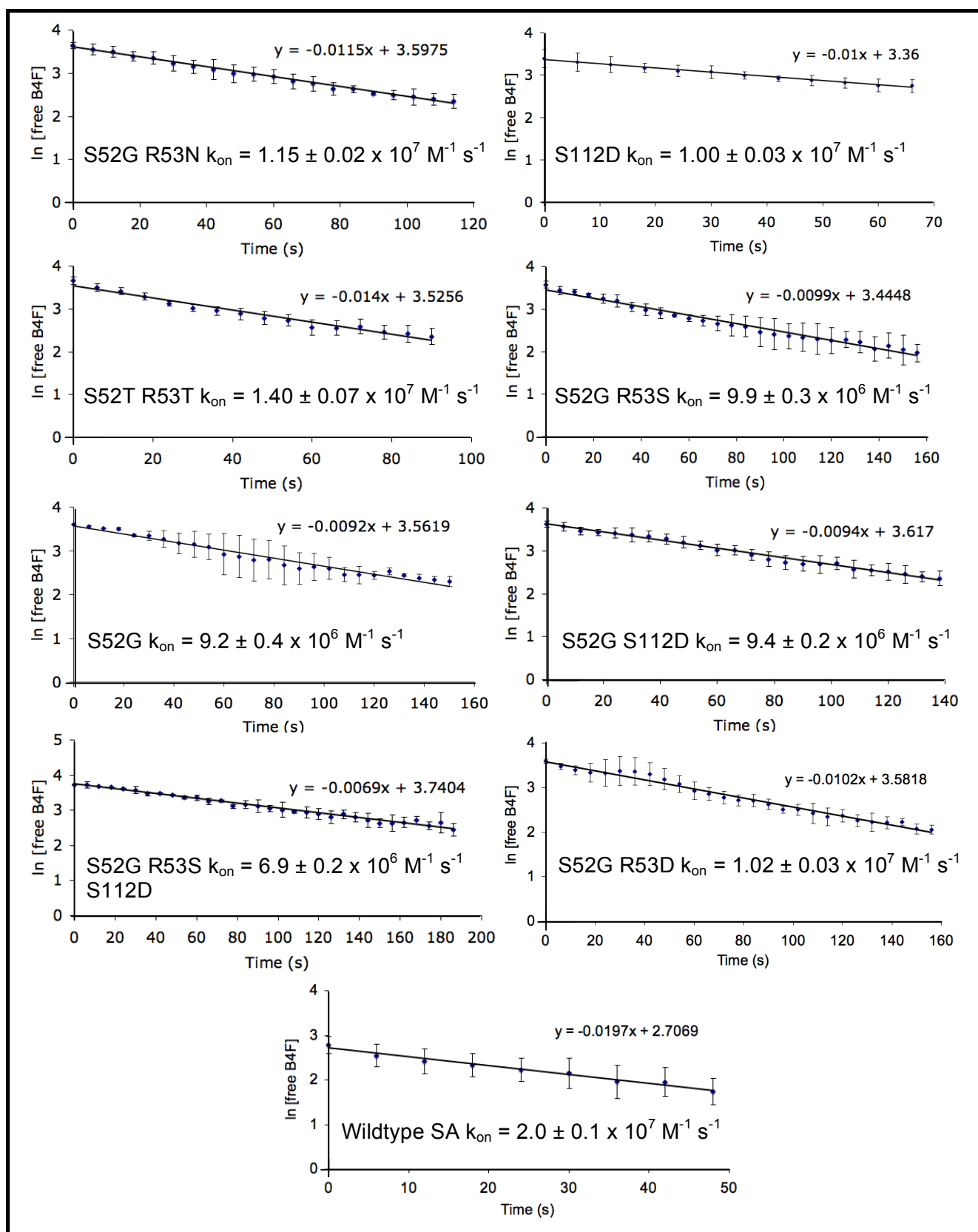


Figure 27. B4F on-rate assay data. Proteins were mixed with B4F in PBS at 25 °C and the rate of fluorescence quenching upon binding measured. Mean of triplicate measurements $\pm$ 1 s.d.

3.3.5 Thermostability

A simple thermostability assay was conducted on every mutant made, together with wildtype SA, in order to obtain an estimate for the temperature at which the dissociation from tetramer to monomer occurs (T_m). After heating the protein at the desired temperature for three minutes, the sample was run on 18 % SDS-PAGE, to visualise the transition from tetramer to monomer as a function of temperature (Figure 28).

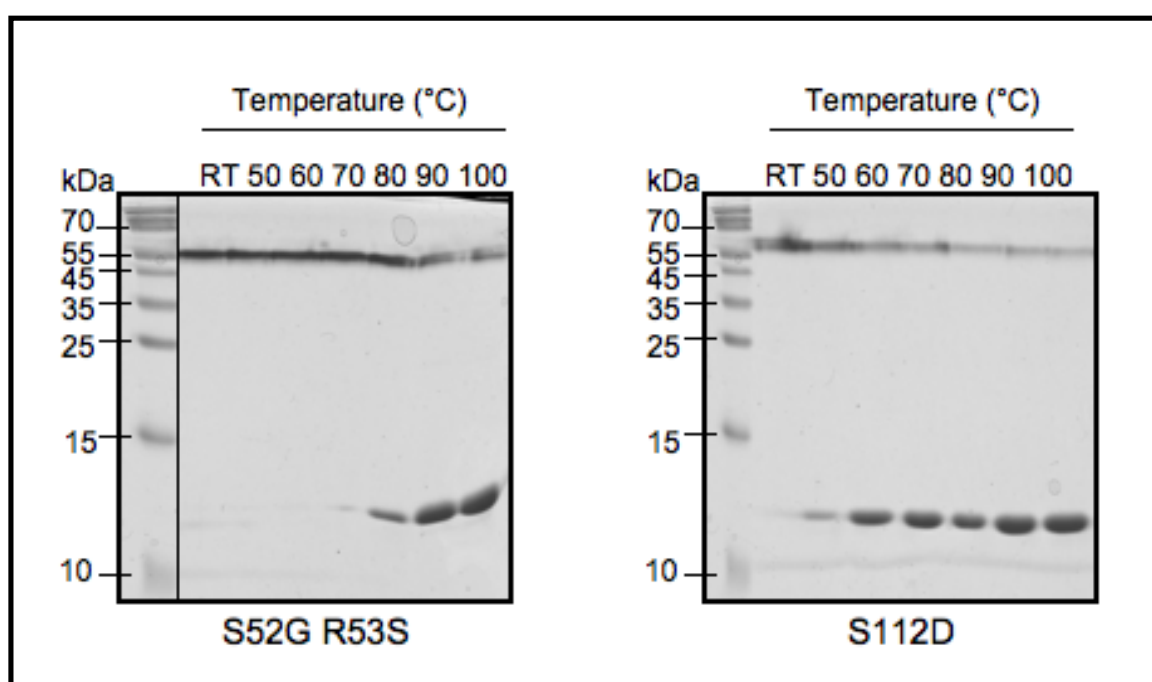


Figure 28. Ascertaining the tetramer-to-monomer transition temperature of wildtype streptavidin and the mutant proteins. Each protein was incubated at the indicated temperature for 3 minutes and then run on 18 % SDS-PAGE. The positive control (C) was boiled in 6X SDS loading buffer prior to loading. Continued on next page.

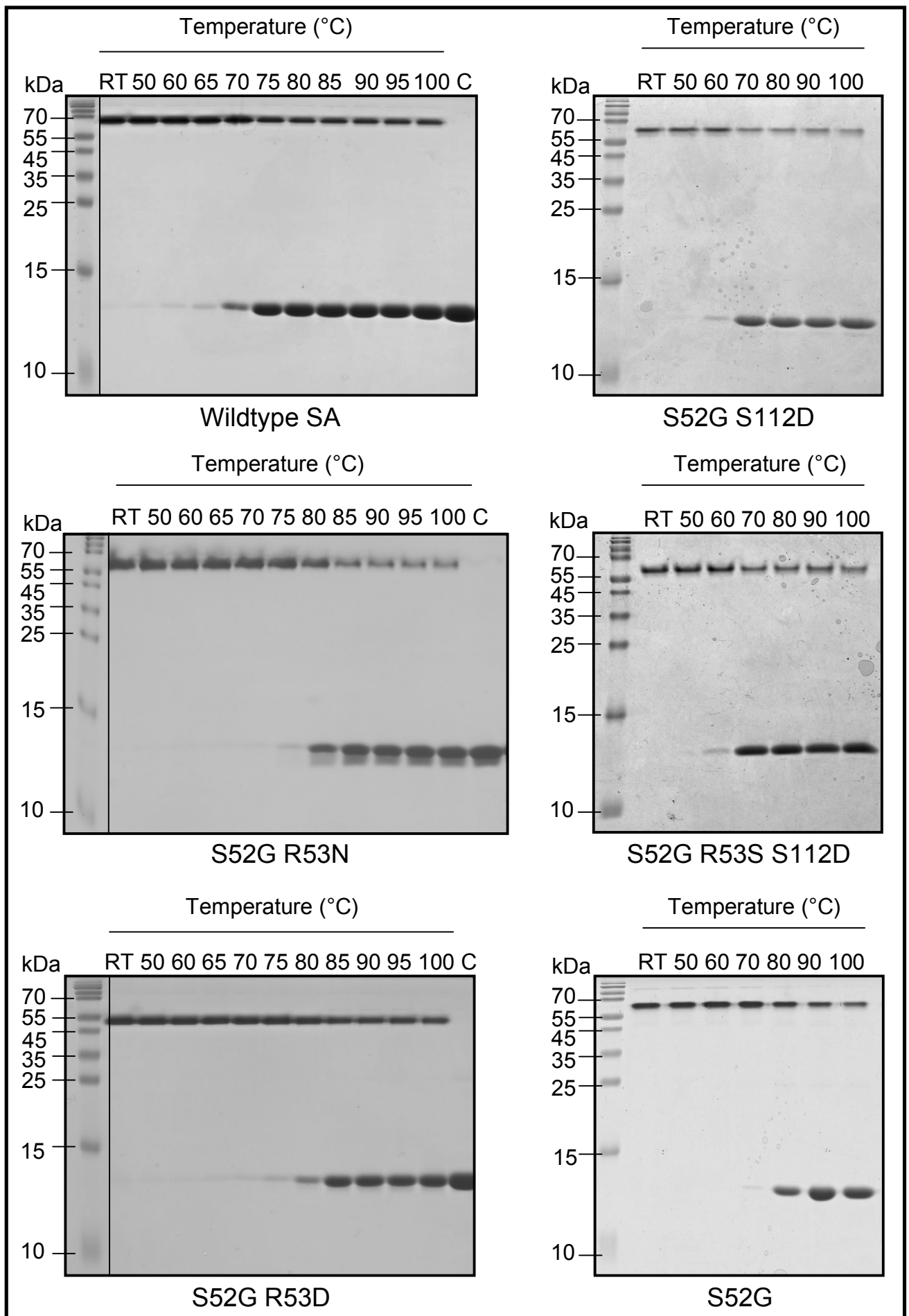


Figure 28 Continued.

By visual inspection of duplicate thermostability gels for SA and each mutant, the temperature at which half the tetramer dissociates to monomer (T_m) can be estimated. This revealed that four of the mutants had increased thermostability (S52G, S52D R53N, S52G R53S and S52G R53D), whilst two mutants had comparable thermostability to wildtype SA (S52G S112D and S52G R53S S112D) and the S112D mutant had reduced thermostability. Although the thermostability gels could have been quantified, only a rapid estimation of the T_m was required here, with detailed quantification of T_m planned for the characterisation of the most promising mutant in Chapter 4.

3.3.6 Summary

In order to decide which specific mutant should be thoroughly characterised further, the yields, thermostability, B4F off-rate and B4F on-rate data were all considered and compared (Table 8).

Protein	Yield (mg/L)	% dissociation B4F after 12 hr	B4F on-rate (M ⁻¹ s ⁻¹)	Relative T _m
Wildtype SA	10.3 ± 4.3	14.1 ± 0.7	2.0 ± 0.1 × 10 ⁷	-
S52G	0.8 ± 0.2	4.1 ± 1.1	9.2 ± 0.4 × 10 ⁶	↑
S112D	0.3 ± 0.1	7.6 ± 1.7	1.00 ± 0.03 × 10 ⁷	↓
S52D R53N	5.6 ± 0.5	7.4 ± 0.9	1.15 ± 0.02 × 10 ⁷	↑
S52G R53S	2.1 ± 0.7	4.3 ± 0.2	9.9 ± 0.3 × 10 ⁶	↑
S52G R53T	0.5 ± 0.1	5.4 ± 0.1	1.40 ± 0.07 × 10 ⁷	ND
S52G S112D	1.4 ± 0.7	1.8 ± 1.5	9.4 ± 0.2 × 10 ⁶	↔
S52G R53S S112D	4.7 ± 1.2	2.0 ± 0.6	6.9 ± 0.2 × 10 ⁶	↔
S52G R53D	9.3 ± 2.7	2.6 ± 2.1	1.02 ± 0.03 × 10 ⁷	↑

Table 8. Summary of initial characterisation data on the eight most promising mutants in comparison to wildtype SA. ↑ indicates increased thermostability, ↓ indicates decreased thermostability and ↔ indicates comparable thermostability to wildtype SA. ND, not determined.

3.4 Discussion

The biotin-4-fluorescein (B4F) off-rate assay was used as the initial screen for biotin-conjugate binding after protein purification. The eight mutants discussed in this chapter had slower B4F off-rates than wildtype SA (14 % dissociation after 12 hours at 37 °C). Two of the mutants, S112D and S52D R53N, showed 7 % dissociation from bound B4F after 12 hours, whilst three of the mutants, S52G R53T, S52G R53S and S52G, showed 4-5 % dissociation. A group of three mutants, S52G S112D, S52G R53S S112D and S52G R53D, had the lowest dissociation rates, of only 2-3 % after 12 hours. Residues Ser-52 and Arg-53 are at the base of L3/4, a flexible loop that is stabilised over the biotin-binding pocket once biotin has bound, contributing to the high-affinity binding (Freitag et al., 1997). These residues were chosen for site-directed mutagenesis as they were found mutated, together with other mutations, in the SA mutants produced in a directed evolution approach to increase desthiobiotin-binding affinity (Levy and Ellington, 2008). Due to their location on L3/4, it was hypothesised that mutations at Ser-52 and Arg-53 could affect the flexibility of L3/4 and exert an effect on the binding stability in this way. Also, as these residues were in the second shell of the hydrogen-bonding network, and do not make direct hydrogen-bonding contact with biotin when it is bound, it was hoped that they would be more amenable to mutagenesis; mutation of the residues in the first shell of the hydrogen-bonding network, that make direct contact with the bound biotin, invariably reduce the binding affinity (Laitinen et al., 2006). The S112D mutation was investigated as Ser-112 is located at the base of the loop connecting β -strands 7 and 8 (L7/8). Mutagenesis of this residue could affect the flexibility and conformation of this

loop, which also contains the conserved residue Trp-120, essential for high-affinity biotin-binding (Chilkoti et al., 1995b; Sano and Cantor, 1995). However, hypotheses were made rather tentatively as it has been shown that affinity can be increased by many small, cumulative mutations, at a distance from the binding site (Midelfort et al., 2004), and protein function can be altered by mutation of purely non-active-site residues (Shimotohno et al., 2001), making prediction of the effects of mutagenesis difficult. Indeed, the large number of weaker-affinity mutants described in Appendix A1, each designed to give more stable binding in theory, is evidence for this.

The production of the eight SA mutants that possessed slower B4F dissociation rates was via an iterative process; if a particular single mutant generated promising data from the B4F off-rate assay, the mutation would be combined with another mutation or mutations, and then this double or triple mutant would have its B4F off-rate measured. In this way, the double mutant S52G S112D was generated in the second round of mutagenesis, after it was found that the single mutants S52G and S112D both had lower B4F off-rates than wildtype SA. In a similar fashion, the S52G R53S S112D mutant was produced and characterised after the S52G R53S and S112D mutants gave promising B4F off-rate data. However, rather unsurprisingly, the majority of the SA mutants that were produced in this study had a faster B4F dissociation rate than the wildtype protein (see appendix A1). This is to be expected, as in a highly optimised system such as the SA-biotin interaction, almost any change is expected to reduce performance, i.e. weaken the interaction and increase the dissociation rate. For this reason, promising mutations from the first round of mutagenesis were also put into mutants that had faster B4F off-

rates than wildtype streptavidin, in order to see if the dissociation rate could be slowed. However, this approach did not decrease the B4F off-rates of the mutants with faster dissociation rates as was hoped. For example, ~ 20 % of bound B4F dissociated from the single mutant A86D after ten hours at 37 °C, whilst only ~ 6 % of bound B4F dissociated from the S112D mutant under the same conditions. When these two single mutations were combined in the second round of mutagenesis to give the double mutant A86D S112D, ~ 50 % of bound B4F had dissociated after ten hours at 37 °C!

All eight of the mutants had a decreased B4F on-rate when compared to wildtype SA. One of the key features of the SA-biotin interaction is its extremely fast on-rate, and so a decreased on-rate was an undesirable outcome of this mutagenesis. However, the slowest B4F on-rate was demonstrated by the mutant S52G R53S S112D, at $6.9 \pm 0.2 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$. This is comparable to the on-rates of many monoclonal antibodies, such as antiluorescein antibodies ($1.3 - 9.8 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$) (Kranz et al., 1982; Midelfort et al., 2004) and anti-dinitrophenol antibodies ($3.4 - 9.5 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$) (James and Tawfik, 2003) and is still an acceptable on-rate for a small molecule-protein association reaction, for which on-rates are typically in the range 0.5 to $5 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$ (Northrup and Erickson, 1992).

The thermostability of each mutant was quickly assessed using a facile SDS-PAGE assay. This revealed that, out of the seven mutants tested, four had moderately increased tetramer thermal stability than wildtype protein. The assay was used to give an approximate initial guide to the thermostability, with the intention of more thoroughly characterising the thermostability of the

most promising stronger-binding mutant in Chapter 4. The temperature at which half the tetramers dissociated to monomers was estimated by eye for SA to be 70 °C, which is comparable to the transition temperature of 75 °C obtained by differential scanning calorimetry (DSC) (Gonzalez et al., 1999). The four mutants S52G, S52D R53N, S52D R53S and S52D R53D had increased thermostability compared to SA, with an estimated T_m of ~ 80 °C. It appears that the S112D mutation has a destabilising effect on the tetramer's thermal stability, as the S112D single mutant had a decreased thermostability (estimated T_m of ~ 55 °C), whilst the S52G S112D and S52G R53S S112D mutants had a comparable thermostability to SA (estimated T_m of ~ 65 °C). This suggests that the mutations S52G, R53N, R53S and R53D gave more favourable charge-charge interactions on the protein surface, which has been shown elsewhere to increase thermostability (Alsop et al., 2003; Gribenko et al., 2009). By the same reasoning, the S112D mutation may give less favourable surface charge-charge interactions, decreasing the thermostability of the protein. Also, proteins with higher thermostability have a tendency towards having shorter, and fewer, exposed loop regions (Thompson and Eisenberg, 1999). The increase in thermostability of the mutants containing the S52G, R53N, R53S and R53D mutations suggests that these mutations could be reducing L3/4 flexibility, whilst the decrease in thermostability of the mutants containing S112D suggests that this mutation may not be having an effect on L7/8 flexibility.

All eight of the mutants with higher-stability biotin-conjugate binding were insoluble when recombinantly expressed in *E. coli*. Whilst SDS-PAGE confirmed all eight mutants had comparable levels of induction in *E. coli*,

some mutants were more amenable to refolding *in vitro* than others, and this was reflected in their typical protein yields. A mutant's yield was borne in mind during the characterisation process, and in the end proved to be a major factor when considering which mutant would be more thoroughly characterised in the following chapters. Three mutants refolded very badly, with yields of < 1 mg/L (S52G, S112D and S52G R53T), suggesting they were more prone to aggregation and precipitation during the refolding process. Four mutants were refolded more successfully, with protein yields between 1 and 6 mg/L. However, strikingly, one mutant, S52D R53D, gave yields comparable to wildtype SA (9.3 mg/L and 10.3 mg/L respectively). Optimisation trials of the rapid dilution method used to refold these mutants were conducted, based on the screening method outlined in Vincentelli et al., 2004. A small selection of refolding buffers was tested, but it was found that PBS better facilitated successful refolding (data not shown). It should be noted, however, that extensive optimisation trials of the refolding method were not conducted on these mutants, and it could be possible to obtain higher yields than those reported in this study by optimisation of refolding factors, such as the rate of decrease of denaturant concentration (rapid dilution versus slow dialysis) during the refolding process, the composition of the refolding buffer and the presence of additives that can aid refolding, such as glycerol (Shimamoto et al., 1998), polyethylene glycol (Cleland et al., 1992) and arginine (Umetsu et al., 2003).

When all the data on the mutants were presented together in Table 6, it was obvious that the S52G R53D mutant was the most promising of the group, having the highest yield (9.3 mg/L), moderately increased thermostability (~

80 °C), extremely low B4F dissociation (3 % after 12 hours at 37 °C) and an acceptable B4F association rate ($1.02 \pm 0.03 \times 10^7 \text{ M}^{-1} \text{ s}^{-1}$). Therefore, it was decided to name this mutant traptavidin (Tr) and carry out further investigation into its biotin-binding properties. Chapter 4 gives details of the further biochemical characterisation of Tr and its comparison to wildtype SA, whilst Chapter 5 presents the crystal structures of both the apo- and biotin-bound forms of Tr.

Chapter 4: Characterisation of Traptavidin

4.1 Introduction

Chapter 3 described the production and characterisation of a group of SA mutants that possessed lower biotin-4-fluorescein (B4F) off-rates than wildtype SA. Taking yield, thermostability, B4F off-rate and B4F on-rate into account, the S52G R53D mutant, named traptavidin (Tr), was deemed the most promising mutant to proceed with. This chapter describes the detailed characterisation of Tr and its comparison to SA and avidin (Av).

Both the B4F and [^3H]biotin on- and off-rates for Tr and SA were measured, to determine the kinetic parameters of these interactions. The B4F binding in particular was examined in greater detail than in Chapter 3, at various pH and over an extended time period. The thermostabilities of Tr and SA tetramers were also quantified, as well as assessing the thermostability of conjugate binding, using a gel-shift assay with biotinylated DNA.

The backgrounds to the novel assays used in this chapter (live cell imaging to assess Tr's non-specific binding, as well as FtsK displacement assays and AFM experiments to assess mechanical stability) are detailed below.

4.1.1 Visualising cell surface IGF1R with the BirA / AP system

The binding specificity of Tr on mammalian cells was determined and compared to that of SA, using the BirA / AP system to visualise cell surface type 1 insulin-like growth factor receptor (IGF1R) (Howarth et al., 2005). IGF1R is a widely expressed transmembrane glycoprotein with tyrosine kinase activity (Adams et al., 2000). The receptor is often found at elevated levels in a variety of human tumours, and hence has become an attractive target for potential cancer therapies (Lopez-Calderero et al., 2010). The BirA / AP system exploits the ability of the *E.coli* enzyme biotin ligase (BirA) to biotinylate a certain lysine in a specific 15 amino acid sequence, called the acceptor peptide (AP) (Beckett et al., 1999). This AP sequence (GLNDIFEAQKIEWHE) can be genetically fused to the protein of interest, in this case IGF1R, to enable biotinylation. In cells that have been transfected with DNA encoding the BirA protein fused to a signal peptide, BirA is expressed, folds in the endoplasmic reticulum and can specifically biotinylate only proteins containing an AP tag; endogenous mammalian proteins are not biotinylated by BirA (de Boer et al., 2003). The biotinylated protein of interest can then be visualised by labelling with SA or Tr conjugated to a reporter molecule such as a fluorescent dye or quantum dot. This imaging system is also discussed in Section 1.4.2.1.

4.1.2 FtsK displacement assays

4.1.2.1 FtsK discovery

The DNA translocase FtsK was first discovered in 1995 in *E. coli* (Begg et al., 1995). A screen of cell division mutants revealed a particular strain that, whilst exhibiting normal behaviour at low temperature, formed long, filamentous cells at 42 °C as a result of a mutation in a temperature-sensitive protein involved in cell division, that was named FtsK (from Filamentous Temperature-Sensitive) (Begg et al., 1995). FtsK has since been found to be highly conserved across almost all bacteria species, due to its key role in coordinating sister chromatid segregation with cell division (Bigot et al., 2007).

4.1.2.2 FtsK cellular role

The coordination of DNA replication, chromosome segregation and cell division in eukaryotes is under spatiotemporal regulation via cell cycle checkpoint mechanisms that are well understood. The details of the process in prokaryotes are less clear, but the pivotal role played by FtsK in linking chromosome segregation to cell division is being actively researched.

When bacteria are about to undergo cell division, a Z-ring forms at the division point at midcell, composed of the tubulin-like protein FtsZ (Bi and Lutkenhaus, 1991). The cell membranes at this division point then invaginate, forming the division septum (Harry et al., 2006). FtsK is among the first proteins to be localised to this septum; its recruitment there is dependent on the FtsK_N domain (see Section 4.1.2.3) and interaction with FtsZ (Wang and Lutkenhaus, 1998). At the septum, FtsK then recruits other proteins, such as

FtsQ, FtsL and FtsI (Chen and Beckwith, 2001), which interact and form a large multi-protein complex called the divisome. FtsK may act to stabilise the divisome, perhaps by providing an anchor point at the septum, or may activate the other proteins in the complex (Draper et al., 1998). Membrane-tethered FtsK at the septum may also act as a DNA pump to remove any remaining DNA from this region prior to cell division (Lau et al., 2003) (Figure 29).

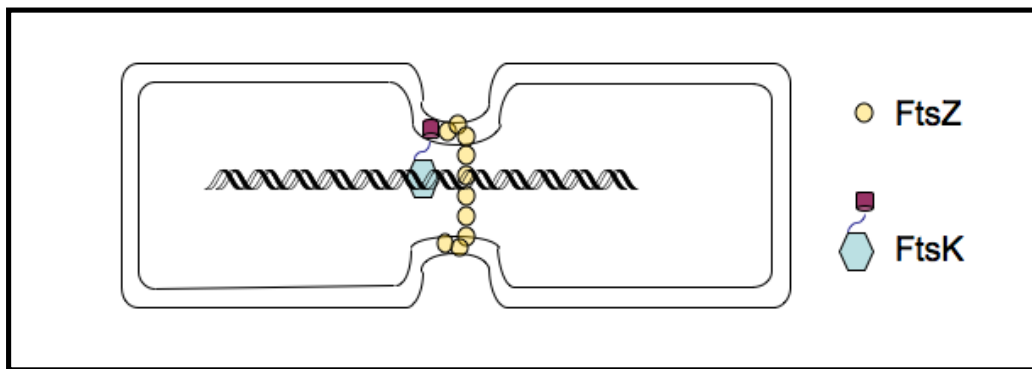


Figure 29. Septum-tethered FtsK is essential for cell division. DNA is pumped through the septum into the daughter cells, removing DNA from the division site.

FtsK is also involved in chromosome segregation. After DNA replication has terminated, sister chromosomes may remain physically linked, either by intercatenation or chromosome dimers. FtsK can disentangle these links, allowing chromosome segregation into daughter cells, prior to cell division (Bigot et al., 2007). Decatenation is brought about by topoisomerase IV (Topo IV), and whilst FtsK is not a requirement for decatenation to occur, Topo IV does interact with FtsK_C (see Section 4.1.2.3), which increases Topo IV activity *in vitro* (Bigot and Mariani, 2010). Chromosome dimers are formed by homologous recombination between sister chromosomes, and bacteria

have a dedicated site-specific recombination system to resolve them. FtsK plays an essential role in the system, which involves the tyrosine recombinases XerC and XerD. The recombination site, *dif*, is located in the termination of replication region, *ter*. FtsK is able to orientate itself on DNA and translocate towards *dif* via its interactions with specific 8 bp sequences, called KOPS (FtsK Orienting / Polarising Sequences; consensus sequence: GGGNAGGG). FtsK preferentially loads onto DNA at KOPS, and interacts with the sequence via its FtsK_C domain (Bigot et al., 2006). This *dif*-biased translocation by septum-tethered FtsK mobilises sister chromosomes to either side of the septum. It also results in FtsK eventually reaching the XerCD/*dif* complex, where it activates the recombinase activity of XerD via its FtsK_C domain (Yates et al., 2006). Despite being able to displace many proteins bound to DNA, such as transcription factors, FtsK is not able to translocate past XerCD/*dif* (Graham et al., 2010). The precise mechanism of XerD activation by FtsK_C remains unknown.

FtsK is a member of the ATPases Associated with various cellular Activities (AAA+) protein superfamily, using ATP hydrolysis to generate the force needed to move its double-stranded DNA (dsDNA) substrate, or, if the DNA is anchored, to translocate along the DNA. FtsK is able to translocate along DNA at speeds between 5 and 6 kb/s, making it the fastest known linear molecular motor (Pease et al., 2005; Saleh et al., 2004). As well as its remarkable speed, FtsK is capable of generating a large force as it translocates, with a stall force of more than 60 pN (Pease et al., 2005). This strength enables FtsK to mobilise DNA in the presence of large counter-

forces, as well as enabling it to displace many of the DNA-bound proteins and RNAs it would encounter during translocation *in vivo* (Graham et al., 2010). In this way FtsK is able to overcome these potential ‘roadblocks’, as well as producing ‘naked’ DNA, stripped of many of its nucleoprotein structures (which may play a role in resetting / clearing epigenetic marks) (Pease et al., 2005).

4.1.2.3 FtsK structure

E. coli FtsK is a 1329 amino acid protein consisting of three distinct domains: FtsK_N, FtsK_L and FtsK_C (Figure 30).

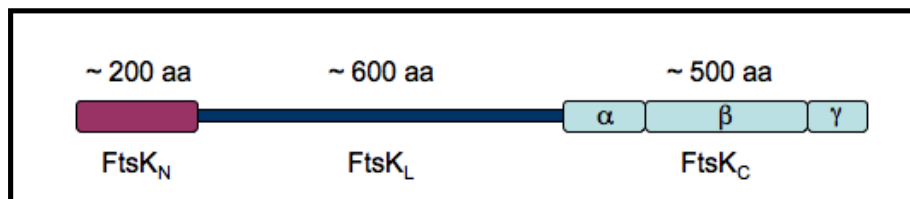


Figure 30. Multidomain structure of FtsK. Approximate sizes of the three domains are shown. The linker domain FtsK_L can vary dramatically in size between species.

The N-terminal domain, FtsK_N, is composed of 217 amino acids in the *E. coli* protein and has poor sequence conservation between species. The structure of this domain is unknown, but it contains transmembrane helices that anchor FtsK in the cell membrane at the division septum. FtsK_N is essential for septum formation in the dividing cell (Draper et al., 1998) and therefore is vital for cell division. FtsK_N is localised into the divisome / septal ring (a multiprotein complex that forms a constrictive ring at midcell, enabling septation to occur) via interactions with FtsZ and other proteins, and can recruit further septum proteins to the divisome (Di Lallo et al., 2003).

Linking the N- and C-termini of the protein is the central linker domain, FtsK_L, which is highly variable in length and sequence, but consists of 600 amino acids in *E. coli*. FtsK_L is rich in proline and glutamine residues, and can form coiled coils, although the exact function of this domain is unclear. FtsK_L is, however, required for successful cell division (Bigot et al., 2004).

The C-terminal domain of FtsK, FtsK_C, consists of 512 aa in the *E. coli* protein and is essential for chromatid segregation, as well as pumping DNA out of the invaginating septum. FtsK_C is itself composed of three subdomains: α , β and γ .

Subunits α and β together form the motor domain, and when crystallised, were found to oligomerise into a hexameric ring, with the α subdomains forming a smaller ring that sits ontop of a larger ring formed by the β subdomains (PDB codes 2IUT and 2IUU) (Massey et al., 2006) (Figure 31). A 30 Å diameter channel is formed at the centre of the hexameric rings, through which dsDNA translocates. The β subdomain possesses ATPase activity and has a RecA-like fold. This structural motif (composed of a central β -sheet flanked by α helices) is found in many ATPases that use the energy of ATP hydrolysis to carry out mechanical work (Ye et al., 2004), and in the β subdomain, is responsible for generating the force needed for DNA translocation. The exact mechanism of translocation through the channel remains unknown, but both the 'rotary inchworm' (Massey et al., 2006) and

'staircase' model (Croizat et al., 2010) have been proposed, involving sequential ATP hydrolysis by adjacent subdomains around the ring.

The γ subdomain has a winged helix-turn-helix structure and is capable of both DNA and protein binding. In this way, it acts as a regulatory domain, binding to KOPS motifs in DNA via helix-3 (Sivanathan et al., 2006) (orientating FtsK on DNA prior to translocation) and binding to the XerD recombinase protein via loop-1 (activating recombination) (Yates et al., 2006).

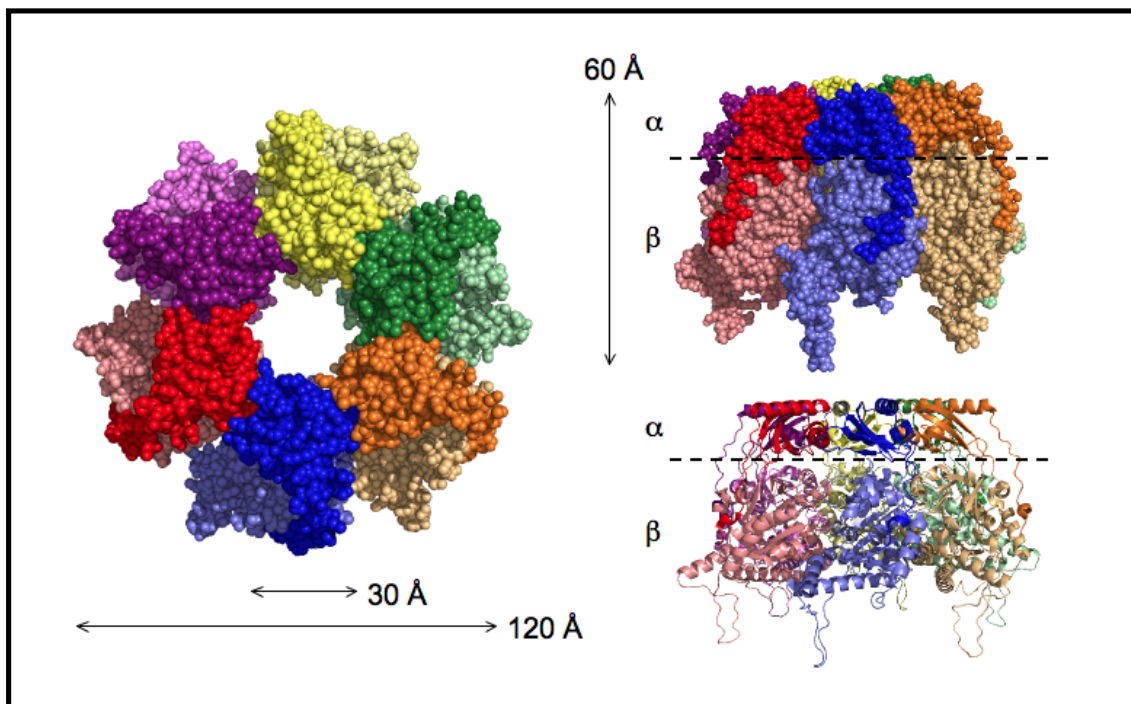


Figure 31. Structure of the motor domain of *P. aeruginosa* FtsK. The α and β subdomains of FtsK_C were crystallised and found to oligomerise to give a hexameric ring, with a central channel 30 Å in diameter through which DNA translocates (Massey et al., 2006). Individual monomers are shown in unique colours, with the α subdomain coloured a darker shade than the β subdomain. The left image shows the top view of the hexamer looking into the channel. A side view of the hexameric motor is shown on the right. The cleft between the α and β subdomains can be clearly seen when presented in cartoon format in the bottom right image. PDB code 2IUU.

4.1.2.4 Constructs used in this study

The FtsK constructs utilised in this study were from the bacteria *E. coli* and *P. aeruginosa* (Figure 32).

The *E. coli* FtsK protein used in this study is the well-studied FtsK_{50C} protein. This construct contains a 50 amino acid region from FtsK_L fused to FtsK_C. *In vitro*, FtsK_C on its own fails to oligomerise efficiently, and needs the addition of the short section of the linker region to be able to hexamerise and form a functioning translocase motor (Aussel et al., 2002).

Experiments were also conducted using *P. aeruginosa* FtsK, PAK4. This construct again contained a short section of FtsK_L fused to the complete FtsK_C domain (Massey et al., 2006).

The FtsK_C domains from the two species have similar folds and a 64 % sequence identity (Crozat and Grainge, 2010).

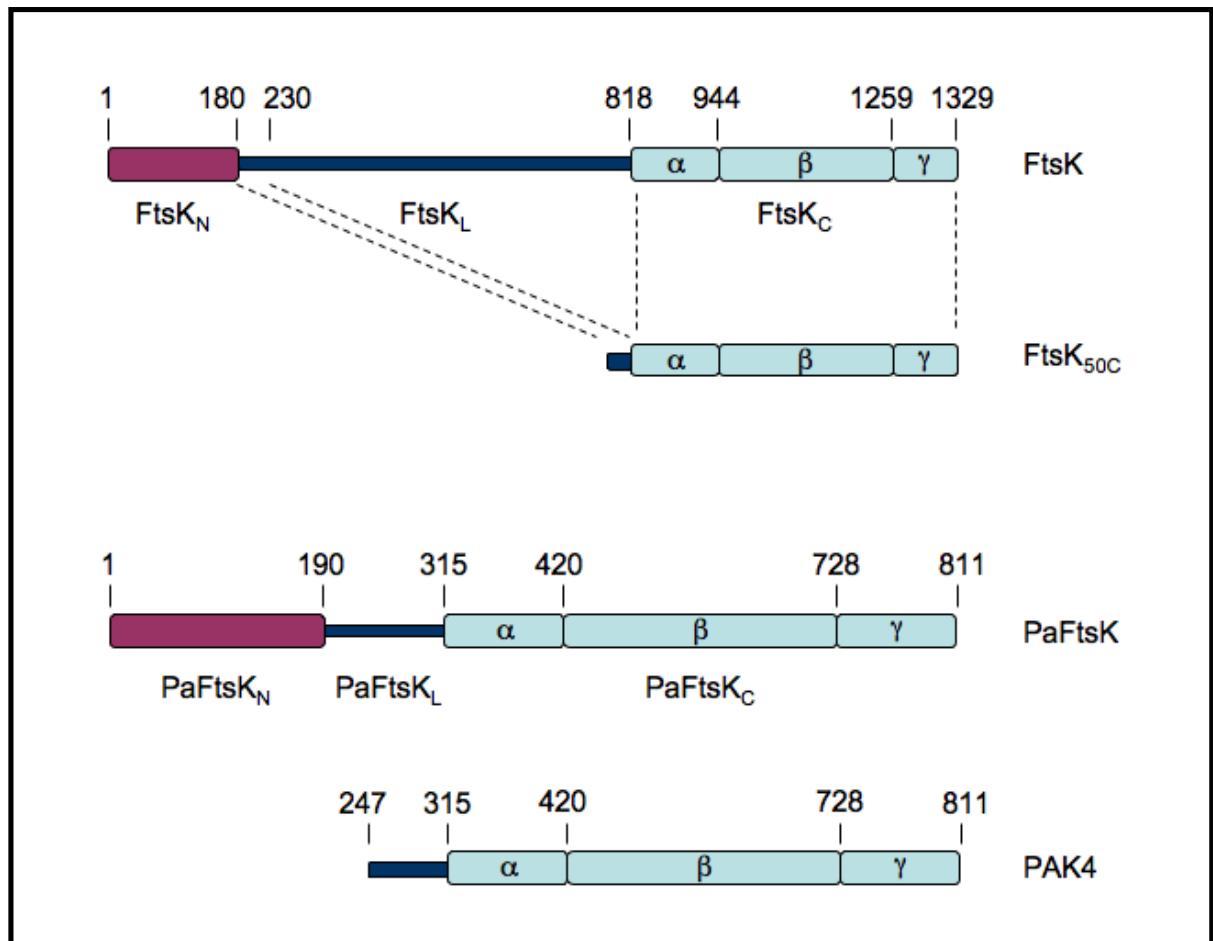


Figure 32. *E. coli* and *P. aeruginosa* FtsK constructs used in this study. *E. coli* FtsK_{50C} contains a 50 amino acid region from the start of the linker domain fused to FtsK_C (Aussel et al., 2002). *P. aeruginosa* PAK4 also contains a region of the linker fused to PaFtsK_C (Massey et al., 2006).

4.1.2.5 Investigating FtsK

FtsK is a tremendously fast and powerful multifunctional molecular motor, capable of both orientated translocation along DNA and activation of recombination, allowing the resolution of chromosome dimers. Although the subject of more than a decade of research, questions still exist regarding this pivotal cell-division protein. The exact mechanism of DNA translocation through the channel, the details of its interaction with XerD that lead to activation of recombination, and the force it is capable of generating during translocation are just a few features of FtsK still to be ascertained (Croizat and Grainge, 2010).

Many different assays have been used to study FtsK. FtsK translocation along DNA has been indirectly followed using magnetic tweezers and optical traps, but as yet, direct imaging of single FtsK molecules remains elusive. The optical trap assay used by Pease et al. to obtain FtsK's stalling force could not allow testing of forces above 60 pN as the dsDNA was distorted and its integrity compromised above this (Pease et al., 2005). Testing FtsK's ability to generate forces in excess of 60 pN requires another approach, notably a facile, bulk *in vitro* biochemical assay that utilises the SA-biotin interaction. In this experimental set-up, the SA protein, bound to a biotinylated DNA substrate, is used to provide a roadblock to translocation. If the force generated during translocation is sufficient to break the stable SA-biotin interaction, SA displacement will occur and be detected in some way (e.g. by a mobility shift on gel electrophoresis). As well as providing another way of demonstrating motor translocation along DNA, the SA roadblock can

give information on the forces generated by motor translocation. Surprisingly, despite the strength of the SA-biotin interaction, certain motor proteins are able to displace SA from biotinylated DNA, including helicases (Morris and Raney, 1999), the mycobacterial motor-nuclease AdnAB (Unciuleac and Shuman, 2010) and FtsK (Chivers et al., 2010). A range of weaker SA mutants can be used to test lower translocation forces (Crozat et al., 2010) and the experimental set-up can be varied to test the coordination and cooperation of multiple motors in overcoming protein roadblocks (Byrd and Raney, 2004).

In this study, Tr and SA were both used as ‘roadblocks’ to the translocation of the molecular motor FtsK (Figure 33). This enabled a comparison of their resistance to molecular motor displacement, as well as shedding light on the force generated by FtsK during translocation. The experimental conditions were also varied to investigate whether multiple FtsK molecules might act together to displace strongly-bound proteins from their translocation path.

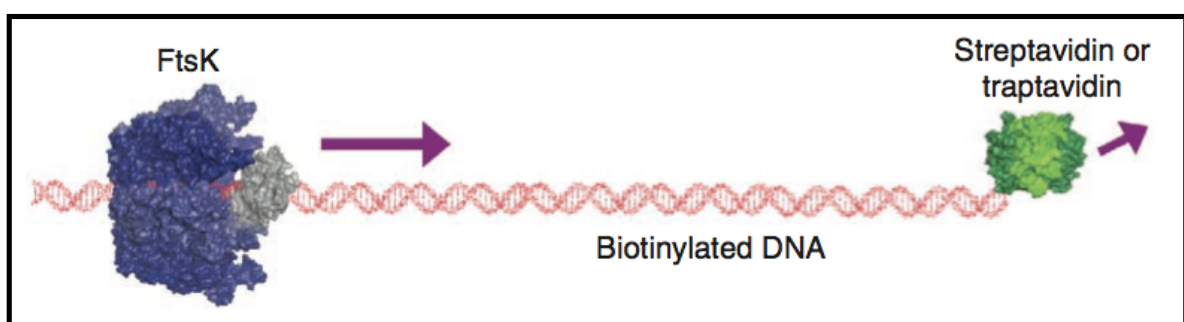


Figure 33. The use of the SA-biotin and Tr-biotin interaction to provide a roadblock to FtsK translocation (Chivers et al., 2010). The α and β subdomains of the FtsK_C domain are shown in purple spacefill format (PDB 2IUU), with the γ subdomain shown in grey spacefill (PDB 2VE9). On addition of ATP, the FtsK protein will translocate along the DNA, until encountering SA or Tr (green spacefill; PDB 1SWE) bound to a biotinylated nucleotide. SA or Tr will only be displaced if the translocation force of FtsK is sufficient to overcome the SA/Tr-biotin interaction.

4.1.3 Atomic force microscopy

The SA-biotin interaction has been used as a paradigm of high-affinity receptor-ligand interaction for many atomic force microscopy (AFM) studies over the years (Florin et al., 1994; Moy et al., 1994; Yuan et al., 2000). These experiments, together with complementary techniques such as biomembrane force probes (BFP) (Merkel et al., 1999) and optical tweezers (Farre et al., 2010), seek to measure the mechanical strength of the interaction at the single-molecule level, by measuring the force needed to break the interaction.

The AFM set-up utilises SA-coated AFM cantilever tips and biotinylated beads; the tip is brought into contact with the beads to allow SA-biotin binding, the cantilever is then retracted and the deflection monitored, and from the force-displacement plot produced, the unbinding / rupture force can be calculated, revealing the strength of the interaction. However, due to the influence of thermal fluctuations when rupturing the bond, repeated rupture measurements of the same interaction under the same experimental conditions can and do give different results, producing a broad spectrum of rupture forces. Typically, the cantilever is retracted (i.e. the interaction is ruptured) hundreds of times during an experiment, producing a histogram of rupture forces, with the peak in this histogram being quoted as the interaction's rupture force.

The AFM experiments can also give information about the dissociation pathway taken by biotin as it is pulled out of SA's binding pocket, and in particular about the location of energy barriers to biotin dissociation along the

pathway. The rupture force measured for an interaction is dependent on the 'loading rate', which is the rate at which force is applied to the interaction. A plot of the rupture force against loading rate over a wide range of loading rates will give information on the number of energy barriers in the dissociation energy landscape, as a certain binding regime (i.e. a certain energy barrier) will create a linear region in the plot. The AFM experiments conducted on the SA-biotin interaction to date have had two linear regions in the rupture force-loading rate plots, corresponding to two energy barriers, whilst the BFP experiments have revealed three linear regions and hence three energy barriers (Merkel et al., 1999; Pierres et al., 2002; Yuan et al., 2000). This discrepancy is a result of the technique being used and is referred to as the SA-biotin paradox (Pincet and Husson, 2005). It is thought to arise from differences in the SA-biotin contact time between the different techniques; when a shorter time period is given for binding before force is applied, the interaction is not able to reach the deepest energy minimum in the pathway, and so this energy barrier may not be observed.

4.2 Materials and methods

4.2.1 Protein production, expression and purification

As described in Chapter 3.

4.2.2 Biotin-4-fluorescein off-rate assays

4.2.2.1 Comparing SA, Tr and avidin B4F off-rates

To allow B4F to bind, 1 μ M SA, Tr or avidin (Sigma-Aldrich, UK) (Av) in 10 μ L PBS was added to 12 nM B4F (Invitrogen, UK) with 0.12 mg/mL bovine serum albumin (BSA) in 170 μ L PBS at pH 7.4 and incubated for 1 hr at 37 °C. 20 μ L PBS or 20 μ L 1 mM biotin in PBS was then added and fluorescence measurements started immediately at 37 °C. Fluorescence measurements of triplicate samples were taken every 10 minutes for 12 hours. Percentage dissociation was calculated as (signal with biotin – signal without biotin)/(signal without quenching – signal without biotin) multiplied by 100. For the signal without quenching, no SA/Tr/Av was added to the B4F. The 'signal without biotin' allowed the fluorescence readings to be corrected for the residual background fluorescence arising from any unbound B4F. Means and standard deviations were calculated from the triplicate measurements. p-Values were calculated using two-tailed Student's *t*-tests from the 6 hr timepoint triplicate data.

4.2.2.2 Troubleshooting the effect of DMSO on fluorescence

To allow B4F to bind to Tr, 1 μ M Tr in 10 μ L PBS was added to 12 nM B4F with 0.12 mg/mL BSA in 170 μ L PBS at pH 7.4 and incubated for 1 hr at 37 °C. 20 μ L PBS, 20 μ L pure DMSO or 20 μ L 1 mM biotin in PBS (diluted from a 100 mM working solution of biotin in 1 % DMSO) was then added and fluorescence measurements immediately started at 37 °C. Fluorescence measurements, in triplicate, were taken every 10 minutes for 70 minutes. Mean raw fluorescence values were plotted with error bars representing $\pm$ 1 standard deviation.

4.2.2.3 Long-term B4F off-rate assay (pH 7.4)

To allow B4F to bind, 1.3 μ M SA or Tr in 10 μ L PBS was added to 5.2 μ M B4F with 1 mg/mL BSA in 5190 μ L PBS and incubated for 4 hr at 37 °C in 15 mL centrifuge tubes (Greiner, UK). The experiment was then started with the addition of 10 μ L PBS or 10 μ L of 52 mM biotin, with the tubes then sealed and placed at 37 °C. Samples (200 μ L) were taken after 0, 0.5, 1, 10, 18, 70 and 92 hours incubation at 37 °C and placed immediately at -80 °C. At the end of the assay, the samples were thawed and their fluorescence measured on a PHERAstar Plus plate-reader with 480 nm excitation and 520 nm emission (BMG LABTECH, UK). Percentage dissociation was calculated as (signal with biotin – signal without biotin)/(signal without quenching – signal without biotin) multiplied by 100. For the signal without quenching, no SA or Tr was added to the B4F. The 'signal without biotin' allowed the fluorescence readings to be corrected for the residual background fluorescence arising

from any unbound B4F. Means and standard deviations were calculated from the fluorescence readings from independent triplicate tubes.

4.2.2.4 B4F off-rate assay (pH 5)

Since fluorescein emission is decreased at low pH, 100 nM SA or Tr was incubated with 12 nM B4F in 100 mM NaCl, 30 mM sodium citrate pH 5.0 for 3 hr at 25 °C, before adding 100 µM biotin and incubating for various times at 37°C. Samples were then placed on ice at the desired timepoint to block further dissociation, adjusted to pH 7.4 with 1 M HEPES pH 8.3, and fluorescence intensity was immediately measured as above. Measurements were done in triplicate, with means and standard deviations calculated. p-values were calculated using two-tailed Student t-tests from the 6 hr timepoint triplicate data.

4.2.3 [³H]biotin assays

4.2.3.1 [³H]biotin off-rate assay

The biotin off-rate from SA and Tr were determined using a method modified from that described previously (Chilkoti, 1995). 10 nM 8,9-[³H]-biotin (PerkinElmer LAS, UK) was incubated for 1 hr at 25 °C with 250 nM SA or Tr in PBS to allow binding. Non-radioactive biotin was then added to a final concentration of 50 µM to initiate dissociation, with incubation at 37 °C. At each time point, the protein-biotin complex was pulled down by incubation with a 50 % slurry of Ni-NTA resin (equilibrated in 300 mM NaCl, 50 mM Tris, 10 mM imidazole pH 7.8) for 1 hr at 25 °C, followed by centrifugation at 800 g for 3 min. 25 µL of supernatant, containing any unbound radioactive biotin, was then added to 1.5 mL Ultima Gold™ scintillation cocktail (PerkinElmer, UK) and counted in a liquid scintillation counter (LS-5000TD, BeckmanCoulter). The average radioactivity of the supernatant at each time point (x) and the radioactivity of the protein-biotin complex before addition of cold biotin (a) were used to determine the first-order dissociation rate constant from the plot of $\ln(\text{fraction bound})$ ($\ln(a-x/a)$) against time. The linear regression and error bars were calculated using the “LINEST” linear least-squares curve-fitting routine in Excel.

4.2.3.2 [³H]biotin on-rate assay

The on-rate of biotin was determined by incubating 250 pM SA or Tr with 1 nM [³H]biotin in PBS at 37 °C. The binding was halted by addition of 50 µM non-radioactive biotin. Protein was pulled down by incubation with 50 % slurry of Ni-NTA resin (equilibrated in 300 mM NaCl, 50 mM Tris, 10 mM imidazole pH 7.8) for 1 hr at 25 °C, followed by centrifugation at 800 g for 3 min. In this time at 25 °C, dissociation of [³H]biotin is negligible (Klumb et al., 1998). 25 µL of supernatant, containing any unbound [³H]biotin, was added to 1.5 mL Ultima Gold™ scintillation cocktail (PerkinElmer, UK) and counted in a liquid scintillation counter (Wallac 1409, PerkinElmer). The radioactivity of the supernatant in cpm at each time point and the total radioactivity present before addition of protein were measured to calculate [free [³H]biotin]. The plot of 1/[free [³H]biotin] against time then enabled the second-order association rate constant to be determined. The linear regression and error bars were calculated using the “LINEST” linear least-squares curve-fitting routine in Excel.

4.2.3.3 Equilibrium dissociation constant calculation

K_d was calculated from k_{off}/k_{on} . The standard deviation was calculated with the formula $(A \pm a)/(B \pm b) = (C \pm c)$, where $c = C \times \sqrt{[(a/A)^2 + (b/B)^2]}$.

4.2.4 Thermostability assays

To determine tetramer thermostability, 3 μ M protein in PBS was heated at the indicated temperature for 3 min followed by cooling to 10 °C in a Bio-Rad DNA Engine® Peltier Thermal Cycler (Bio-Rad, UK) and then immediately placed on ice (Bayer et al., 1996). Samples were then run on 18% SDS-PAGE. The 100% monomer positive control was mixed with SDS loading buffer prior to heating at 95 °C for 3 min. A ChemiDoc XRS imager and QuantityOne 4.6 software (Bio-Rad, UK) were used to quantify the band intensities.

To determine thermostability of biotin conjugate binding, 5 μ M SA or Tr in PBS was incubated with 21 nM monobiotinylated DNA in a volume of 4 μ L for 30 min at 25 °C. Samples were made up to 10 μ L with a final concentration of 100 μ M biotin, 20 mM Tris acetate, 1 mM DTT, 2 mM magnesium acetate and 20 mM potassium glutamate pH 7.5 and incubated for 5 min at 25 °C, before heating at the indicated temperature for 3 min in a Bio-Rad DNA Engine® Peltier Thermal Cycler. After cooling to 10 °C, samples were then run on a 1.5 % agarose gel at 6 V/cm in TAE buffer (40 mM Tris acetate, 1 mM EDTA pH 8.2) for 45 min at 25 °C. Ethidium bromide-stained DNA was visualised and quantified on a ChemiDoc XRS imager using QuantityOne 4.6 software. Percentages were defined as the intensity of the band for free DNA divided by the summed intensities of the bands for free and bound DNA, multiplied by 100. The 439 bp monobiotinylated DNA was prepared by PCR using the primers Fts1 and the internally biotinylated primer bioFts2 (Eurofins) and

plasmid pJEG41-N1, a derivative of pUC18 containing a lambda phage insert (see Table 9).

4.2.5 Cell imaging with the BirA / AP system

4.2.5.1 Cell culture

COS7 cells were grown in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10 % Fetal Calf Serum, 50 U/mL penicillin and 50 µg/mL streptomycin at 37 °C in 5 % CO₂.

4.2.5.2 Alexa Fluor® 555 labelling of SA and Tr

A ten-fold molar excess of Alexa Fluor® 555 carboxylic acid, succinimidyl ester (Invitrogen, UK) was added to solutions of SA and Tr in PBS with 0.1 M sodium bicarbonate pH 8.3 and incubated at room temperature for 3 hours. 555-conjugated protein was then separated from unconjugated dye by size exclusion chromatography on a G-25 Sephadex column, eluting with PBS, and then dialysed three times against PBS. The degree of labelling of the conjugate was determined according to the manufacturer's instructions, using the absorbance at 280nm and 555nm.

4.2.5.3 Plasmid construction

The AP tag was fused to the insulin-like growth factor-1 receptor (IGF1R) by insertion of the AP sequence, along with a 6 amino acid spacer sequence, after the IGF1R signal sequence and before the start of the sequence

encoding the N terminus of mature IGF1R. This AP-IGF1R construct was produced from pcDNA3 containing human IGF1R (a gift from V. Macaulay, University of Oxford) by amplifying two fragments: the first fragment with primers IGFA and IGFB and the second fragment with primers IGFC and IGFD (Table 9). The fragments were joined by overlap extension PCR, digested with NheI and NotI and then ligated into pcDNA3.1. BirA-ER (*E.coli* biotin ligase targeted to the endoplasmic reticulum) and pECFP-H2B (human histone H2B fused to enhanced CFP; used as a cotransfection marker) have been described elsewhere (Howarth et al., 2008).

4.2.5.4 Transfection and biotinylation

COS7 cells were transfected using Lipofectamine 2000 (Invitrogen, UK) following manufacturer's instructions using 0.25 µg AP-IGF1R, 0.2 µg BirA-ER and 0.05 µg H2B-ECFP per well in a 48-well plate. Cells were incubated with 10 µM biotin overnight for optimum biotinylation by BirA-ER (Chivers et al., 2010). The next day, cells were washed three times in PBS with 5 mM MgCl₂ (PBS/Mg) and then kept at 4 °C for imaging.

4.2.5.5 Cell imaging

To visualise biotinylated AP-IGF1R, cells were incubated for 15 min in PBS/Mg with 1 % dialysed BSA and 0.4 µM Alexa Fluor 555-conjugated Tr or SA. For pre-blocking, 50 µM biotin was added to the fluorescent Tr-555 sample 5 min before adding to cells. Cells were washed with PBS/Mg three times before imaging live. Imaging was conducted using a wide-field DeltaVision Core fluorescent microscope (AppliedPrecision) with a 40X oil-

immersion lens. ECFP (436DF20 excitation, 480DF40 emission, Chroma 86002v1 dichroic) and Alexa Fluor 555 (540D420 excitation, 600DF50 emission, Chroma 84100bs polychroic) images were collected and then analysed using softWoRx 3.6.2 software. Exposure times were typically 0.1 - 0.5 s and fluorescence images were background-corrected. Identical conditions were used for preparation, imaging and analysis for individual samples in the same experiment.

4.2.6 FtsK displacement assays

4.2.6.1 Displacement with *Pseudomonas aeruginosa* FtsK

A soluble fragment of *P. aeruginosa* FtsK, PAK4, was utilised in these experiments (a kind gift of J. Graham, University of Oxford). PAK4 contains the C-terminal 578 residues of FtsK, including the α , β and γ domains, and a His₆ tag (Massey et al., 2006). The protein was recombinantly expressed in *E. coli* and purified by sequential ammonium sulfate precipitation, nickel affinity chromatography and heparin affinity chromatography, as previously described (Sivanathan et al., 2006).

The translocation template was a monobiotinylated 439 bp DNA fragment, generated by PCR using the primers Fts1 and the internally biotinylated primer bioFts2 (Eurofins) from plasmid pJEG41-P1 (see Table 9). This DNA fragment contained two KOPS loading sites (GGGCAGGGGGGCAGGG) positioned such that upon addition of ATP, it would take FtsK loaded at KOPS ~ 0.5 s to translocate to the SA or Tr bound to the internal biotinylated

thymine (Graham et al., 2010). All assays were performed at 25 °C in 20 mM Tris acetate, 2 mM magnesium acetate, 20 mM potassium glutamate and 1 mM DTT (pH 7.5). After PAK4 translocation had occurred and the reactions had been stopped (details below), samples were run on 1.5 % agarose gel electrophoresis in TAE at 6 V/cm for 45 minutes at 25 °C. Ethidium bromide-stained DNA was visualised and quantified on a ChemiDoc XRS imager. The percentage of free DNA in each lane was calculated using the formula: $\text{percentage free} = (\text{intensity of band for free DNA}) / (\text{intensity of band for free DNA} + \text{intensity of band for bound DNA})$ multiplied by 100. The percentage of DNA with displaced SA or Tr was then calculated for each lane using the formula: $\text{percentage displaced} = (\% \text{ free} - \% \text{ free for negative control}) / (100 - \% \text{ free for negative control})$, multiplied by 100. Negative values for percentage displaced were reported as zero.

4.2.6.1.1 Initial PAK4 displacement assay

To allow binding, 16 nM DNA fragment was incubated with 2 µM SA or Tr for 10 minutes, followed by 100 µM biotin to block free biotin binding sites. 0.5 µM PAK4 was then added and allowed to load onto the DNA, at KOPS, for 5 minutes. 2 mM ATP was added to start the translocation. After 3 minutes of PAK4 translocation, the reaction was stopped by adding 1 µL of 0.1 % SDS with 200 mM EDTA (pH 8) and incubated for 15 minutes to allow PAK4 to dissociate from the DNA. The samples were then mixed with 6X gel loading buffer (1.2 M sucrose, 0.75 mM bromophenol blue, 0.93 mM xylene cyanol FF), run on 1.5 % agarose gel electrophoresis and imaged as above.

4.2.6.1.2 Investigating effect of PAK4 translocation time

To allow binding, 16 nM DNA fragment was incubated with 2 μ M SA or Tr for 10 minutes, followed by 100 μ M biotin to block free biotin binding sites. 0.5 μ M PAK4 was then added and allowed to load onto the DNA, at KOPS, for 5 minutes. 2 mM ATP was added to start the translocation. After 0, 2, 10, 30, 60 and 300 s of PAK4 translocation, the reaction was stopped by adding 1 μ L of 0.1 % SDS with 200 mM EDTA (pH 8) and incubated for 15 minutes to allow PAK4 to dissociate from the DNA. The samples were then mixed with 6X gel loading buffer (1.2 M sucrose, 0.75 mM bromophenol blue, 0.93 mM xylene cyanol FF), run on 1.5 % agarose gel electrophoresis and imaged as above.

4.2.6.1.3 Investigating effect of PAK4 concentration

To allow binding, 16 nM DNA fragment was incubated with 2 μ M SA or Tr for 10 minutes, followed by 100 μ M biotin to block free biotin binding sites. 0, 0.1, 0.2, 0.5, 1 and 2 μ M PAK4 was then added and allowed to load onto the DNA, at KOPS, for 5 minutes. 2 mM ATP was added to start the translocation. After 3 minutes of PAK4 translocation, the reaction was stopped by adding 1 μ L of 0.1 % SDS with 200 mM EDTA (pH 8) and incubated for 15 minutes to allow PAK4 to dissociate from the DNA. The samples were then mixed with 6X gel loading buffer (1.2 M sucrose, 0.75 mM bromophenol blue, 0.93 mM xylene cyanol FF), run on 1.5 % agarose gel electrophoresis and imaged as above.

4.2.6.2 Displacement with *Escherichia coli* FtsK

A soluble fragment of *E. coli* FtsK, FtsK_{50C}, was also utilised in this study (a kind gift of E. Crozat, University of Oxford). FtsK_{50C} contains the α , β and γ domains of FtsK_C, as well as a 50 aa region of FtsK_L that is necessary for the formation of the hexameric motor. FtsK_{50C} was overexpressed in *E. coli* and purified as described previously (Lowe et al., 2008).

The translocation template was a monobiotinylated 598 bp DNA fragment, generated by PCR using the primers Fts1 and the terminally biotinylated primer bioFts3 (Sigma-Genosys) from plasmid pJEG41-P1 (see Table 9). This DNA fragment contained two KOPS loading sites (GGGCAGGGGGGCAGGG) 226 bp from the biotinylated end of the fragment. All assays were performed at 25 °C in 25 mM Tris, 10 mM MgCl₂ (pH 7.5).

To allow binding, 5.9 nM of the DNA translocation template was incubated with 0.5 μ M SA or Tr for 15 minutes, followed by 100 μ M biotin to block free biotin binding sites. 1 μ M FtsK_{50C} was then added and allowed to load onto the DNA, at KOPS, for 5 minutes. 2.5 mM ATP was added to start the translocation. After 2 minutes of FtsK_{50C} translocation, the reaction was stopped with a final concentration of 0.1 % SDS with 20 mM EDTA (pH 8) and incubated for 20 minutes to allow FtsK_{50C} to dissociate from the DNA. The samples were then mixed with 10X gel loading buffer (250 mM Tris, 20 mM EDTA, 50 % glycerol, 2.5 % bromophenol blue pH 7.5) before running on a 1.5 % agarose gel in TBE buffer (89 mM Tris, 89 mM boric acid, 2 mM EDTA

pH 8.3) at 10 V/cm for 2 hours at 25 °C. The gel was stained with SYBR Green (Invitrogen) for 2 hours, washed in double-distilled water for 30 minutes, and then imaged and quantified using a Fuji FLA3000 scanner and Image Gauge Software (Fuji). The percentage of free DNA in each lane was calculated using: $(\text{intensity of band for free DNA}) / (\text{intensity of band for free DNA} + \text{intensity of band for bound DNA})$ multiplied by 100. The percentage of DNA with displaced SA or Tr was then calculated for each lane using: $(\% \text{ free} - \% \text{ free for negative control}) / (100 - \% \text{ free for negative control})$, multiplied by 100. Negative values for percentage displaced were reported as zero.

4.2.7 Atomic force microscopy

This work was kindly carried out by collaborators, Calvin Chu and Vincent Moy, at the University of Miami.

AFM cantilevers (Veeco MLCT-AUHW) with a nominal spring constant of 0.01 N/m were used in all measurements and calibrated by the thermal fluctuation method (Hutter and Bechhoefer, 1993). Cantilevers were briefly washed in acetone and then UV-irradiated for 15 min. Cantilevers were then briefly dipped in 0.1 M NaHCO₃ pH 9, air-dried, and coated in 0.5 mg/mL biotin-amidocaproyl-bovine serum albumin (biotin-BSA) (Sigma) overnight at 4 °C in a humidified chamber. Cantilevers were washed three times in PBS and then coated with 0.01-0.3 mg/mL SA or Tr for 15 min at 25 °C. 35 mm² dishes were incubated with 0.1 M NaHCO₃ pH 9 followed by 0.5 mg/mL SA (from *Streptomyces avidinii*, Pierce) in 0.1 M NaHCO₃ pH 9 overnight in a humidified chamber at 4 °C. Biotinylated agarose beads (Sigma) were then aspirated onto the dish. Measurements were conducted at 25 °C with a custom built AFM (Rico and Moy, 2007). Functionalised cantilevers were pressed against the agarose beads and the cantilever was retracted from the bead at 374-532 nm/s. To ensure that in most cases no or only 1 bond formed between the cantilever and the bead, the beads were indented with a minimal amount of force of 25-50 pN and 2.5 µg/mL biotin-BSA was added to titrate the binding sites on the cantilever, to achieve 30-40 % adhesion frequency. Single molecule unbinding forces were determined by calculating the force in the retraction trace at the initial point of rupture, using Igor Pro 6.04

(Wavemetrics Inc.). Observed forces were corrected for the hydrodynamic drag force (Alcaraz et al., 2002).

The loading rate was determined as the slope of the force versus time plot, before the rupturing of the single molecule bond. The best fit of the unbinding force to the loading rate was calculated from the Bell model:

$$F = k_B T / \gamma \cdot \ln(\gamma r_f / 2k^0 k_B T)$$

(where F is unbinding force, r_f is loading rate, k_B is the Boltzmann constant, T is absolute temperature, k^0 is the dissociation rate in the absence of force, and γ is the position of the transition state) by minimizing the chi-square statistic in Igor Pro 5.05. Deriving the Bell model parameters (± 1 s.e.m.), k^0 is $0.88 \pm 0.16 \text{ s}^{-1}$ for SA and $0.51 \pm 0.04 \text{ s}^{-1}$ for Tr, while γ is $2.80 \pm 0.12 \text{ \AA}$ for SA and $2.60 \pm 0.06 \text{ \AA}$ for Tr. The similar γ values suggest that there is no substantial difference between the two proteins in the distance of the potential barrier from the centre of the binding pocket, as biotin is pulled out. This is consistent with each protein having the same pathway for biotin dissociation, Tr having the higher energy barrier to mechanical dissociation. The 95 % confidence bands were determined from a two-tailed Student t-test, with 8 (Tr) or 4 (SA) degrees of freedom. To compare unbinding over this range of loading rates, significance was determined by a two-way ANOVA without replication (SA $n = 400$, Tr $n = 562$).

The broad distribution of binding strengths relates to the significance of thermal fluctuations to traversing the activation barrier (Evans and Ritchie, 1997); significance was determined with the Mann-Whitney test (streptavidin $n = 265$, traptavidin $n = 249$).

In order to plot the relationship between dissociation rate and force, we calculated the force-dependent dissociation rate k :

$$k = k^0 \cdot \exp(\gamma f / k_B T)$$

where k^0 is the dissociation rate in the absence of force, γ is the position of the transition state, f is the pulling force, k_B is the Boltzmann constant, and T is absolute temperature (Evans and Ritchie, 1997).

Name	Sequence (5' – 3')
Fts1	CGGAGACGGTCACAGCTTG
bioFts2	CGGCTCGTA[biotin-dT]GTTGTGTGG
bioFts3	[biotin]CGGCTCGTATGTTGTGTGG
IGFA	CAACGCTAGCGCCGCCACCATG
IGFB	CACTCGATCTTCTGGGCCTCGAAGATATCGTTCAGGCCGCCAAT
	CTCTCCACTCGTCGGCCAGAG
IGFC	GAGGCCCAAGATCGAGTGGCACGAGGGCAGTGAAGGATCT
	GGCGAAATCTGCGGGCCAGGC
IGFD	GATTGCGGCCGCTCAGCAGGTC

Table 9. DNA oligonucleotides used in this study

4.3 Results

4.3.1 Biotin-4-fluorescein off-rate assays

In order to appreciate the extremely low B4F off-rate from Tr at 37 °C, a B4F off-rate assay at pH 7.4 was conducted, comparing SA, Tr and the eukaryotic biotin-binding protein avidin (Av) (Figure 34). Despite Av having a higher affinity for free, unconjugated biotin than SA, SA binds biotin-conjugates more strongly than Av (Pazy et al., 2002), and specifically, has a dramatically lower off-rate from B4F than Av (~ 80 % of bound B4F dissociates from Av after 12 hours at 37 °C, whilst in this time only ~ 14 % B4F dissociates from SA). However, the B4F off-rate from Tr is even more striking, with an initial ~ 2 % dissociation at time 0, followed by very little further dissociation over the subsequent 12 hours. This significantly slower dissociation from Tr ($P = 0.0008$) was extremely promising and led to further characterisation of the mutant's binding stability.

The percentage quenching of B4F fluorescence obtained at the beginning of the assay before addition of excess free biotin is shown in Table 10. On binding B4F, SA and Tr both gave 90 % quenching of fluorescence, whilst Av gave ~ 70 % quenching.

Protein	Quenching (%)
Streptavidin	90.3 ± 0.3
Avidin	66.6 ± 1.1
Traptavidin	90.4 ± 0.5

Table 10. Percentage quenching of B4F fluorescence by SA, Av and Tr binding. Means of triplicate measurements shown ± 1 s.d.

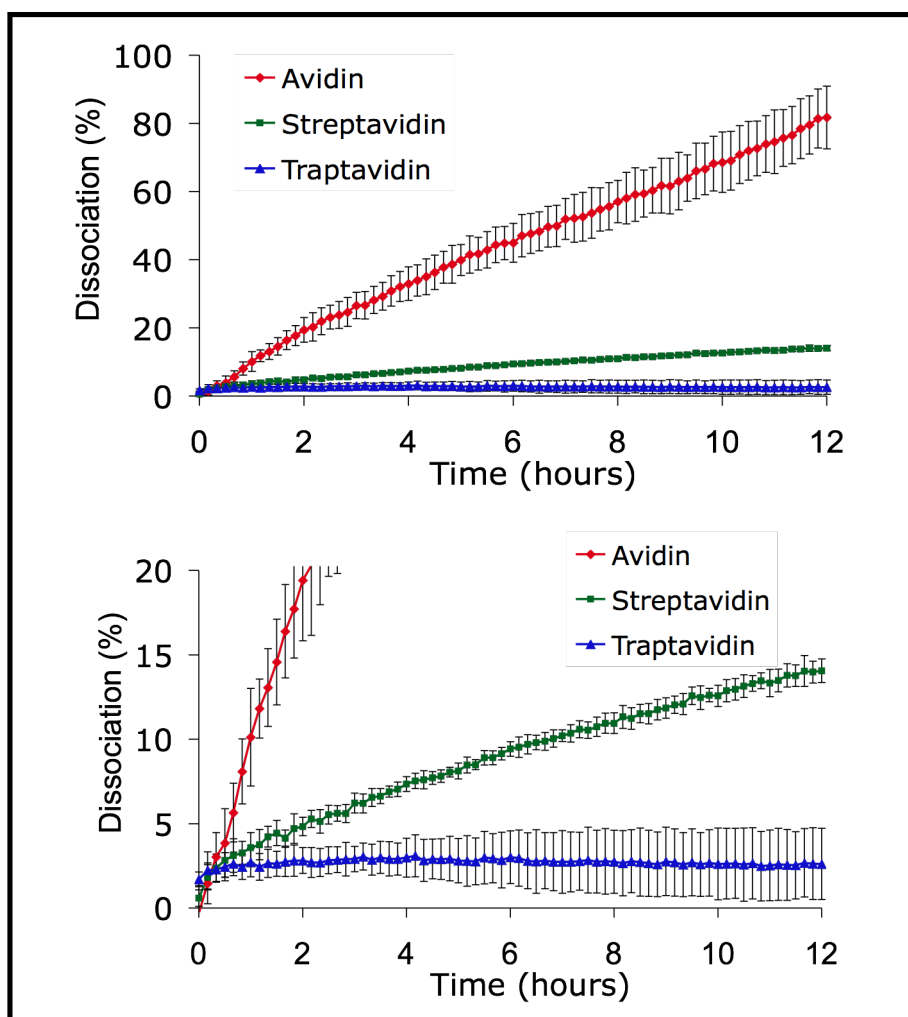


Figure 34. Traptavidin has the lowest B4F off-rate, when compared to SA and Av. Av has a lower affinity for biotin-conjugates than both SA and Tr (upper panel). Although a much better biotin-conjugate binder than Av, SA's B4F off-rate has been improved upon, in the mutant Tr (lower panel). Assay conducted at 37 °C in PBS with 0.12 mg/mL BSA. Mean values of triplicate experiments plotted, with error bars representing ± 1 s.d.

To ensure the initial $\sim 2\%$ B4F dissociation from Tr at time 0 was not an artifact of the experimental set-up, the effect of a low concentration of DMSO (0.1 % final concentration), which was present in the biotin solution, on fluorescence was measured (Figure 35). After incubating B4F with SA for 1 hour at 37°C , the addition of biotin caused an initial decrease, followed by an increase in fluorescence levels, due to the dissociation of B4F from SA, as expected. However, the addition of either PBS or DMSO also caused an initial decrease in fluorescence due to dilution effects, followed by a steady fluorescence level over the next hour. Therefore, the initial jump in B4F dissociation seen in Figure 34 is not a result of a small amount of DMSO causing an increase in fluorescence levels.

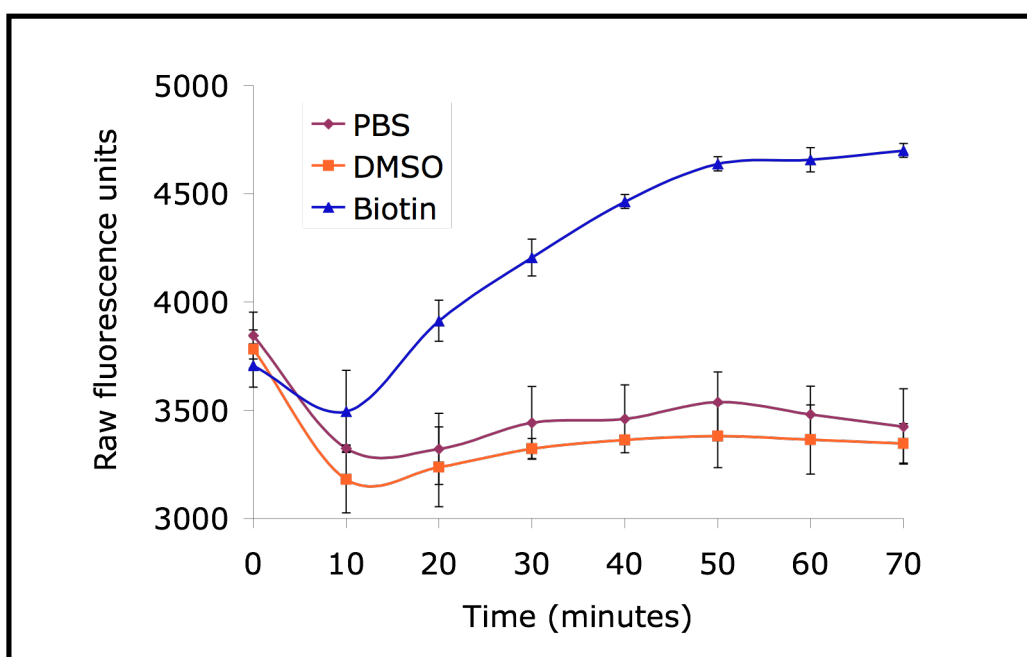


Figure 35. The initial 2 % B4F dissociation from Tr at time 0 is not due to the presence of DMSO in the biotin sample. Mean of triplicate experiments plotted ± 1 s.d. Some error bars are too small to be visible.

In order to assess B4F binding stability over a longer time period at 37 °C, an extended B4F off-rate assay at pH 7.4 was conducted with both SA and Tr (Figure 36). SA steadily dissociated over time at 37 °C, whilst most B4F remained bound to Tr, reflecting Tr's increased binding stability.

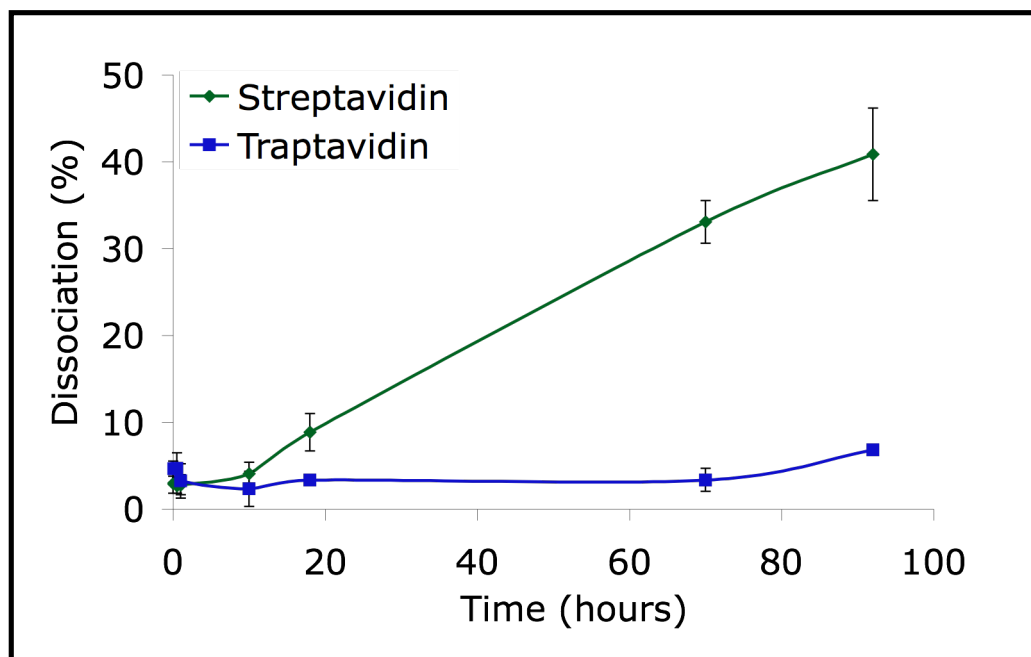


Figure 36. Long-term B4F off-rate assay at pH 7.4. Most B4F remains stably bound to Tr over a longer time-period at 37 °C, whilst B4F bound to wildtype SA steadily dissociates. Mean of triplicate measurements $\pm$ 1 s.d. Some error bars are too small to be visible.

In order to assess the stability of B4F binding at a lower pH, the B4F off-rate assay at pH 7.4 was adapted in order to measure the dissociation rate at pH 5 (Figure 37). At this lower pH, B4F dissociation was faster for both SA and Tr than at neutral pH, but the pattern of B4F dissociation was the same, i.e. Tr having significantly slower dissociation than SA ($P = 0.001$).

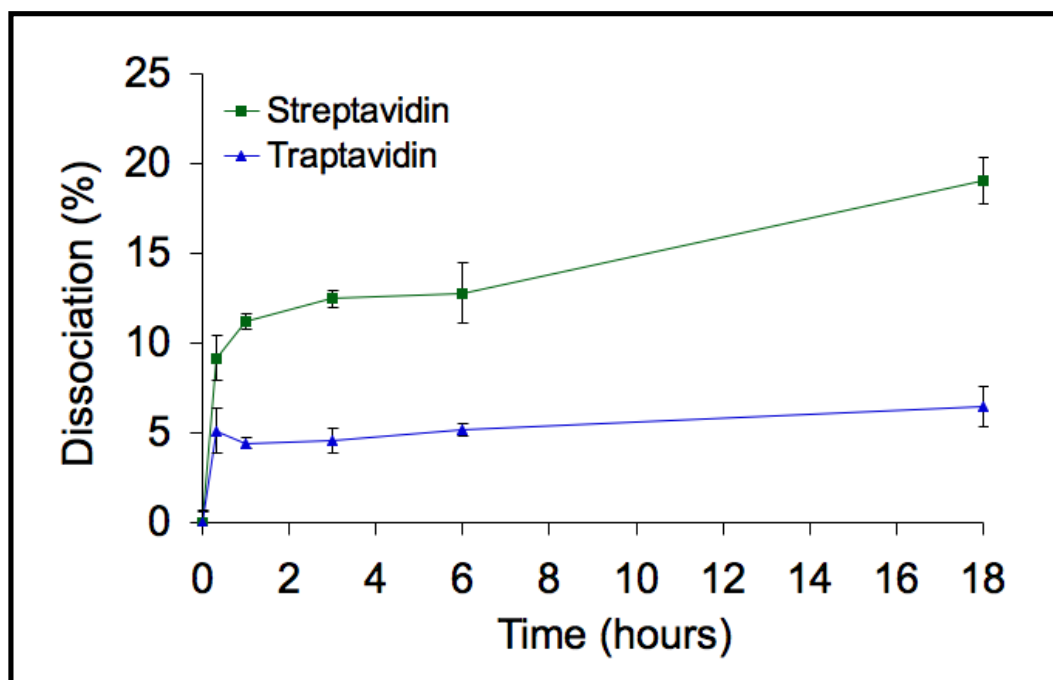


Figure 37. Traptavidin has a lower B4F off-rate at pH 5. Mean of triplicate measurements ± 1 s.d. Assay conducted at 37 °C with data collected by M. Howarth.

4.3.2 [^3H]biotin assays

After the B4F off-rate assays revealed Tr's increased binding stability to this biotin-conjugate, it was decided to measure Tr's affinity for free biotin, in order to ascertain whether the mutations in Tr affected its ability to bind a specific biotin-conjugate, or its general biotin-binding ability. The [^3H]biotin off-rate assay, conducted at 37 °C and pH 7.4, revealed Tr had more than a ten-fold lower off-rate than SA (Figure 38); $4.2 \pm 0.5 \times 10^{-6} \text{ s}^{-1}$ for Tr compared to $6.8 \pm 0.3 \times 10^{-5} \text{ s}^{-1}$ for SA. This suggested that Tr's decreased off-rate was not specific to a particular biotin-conjugate, but instead may be a general property of Tr's binding ability.

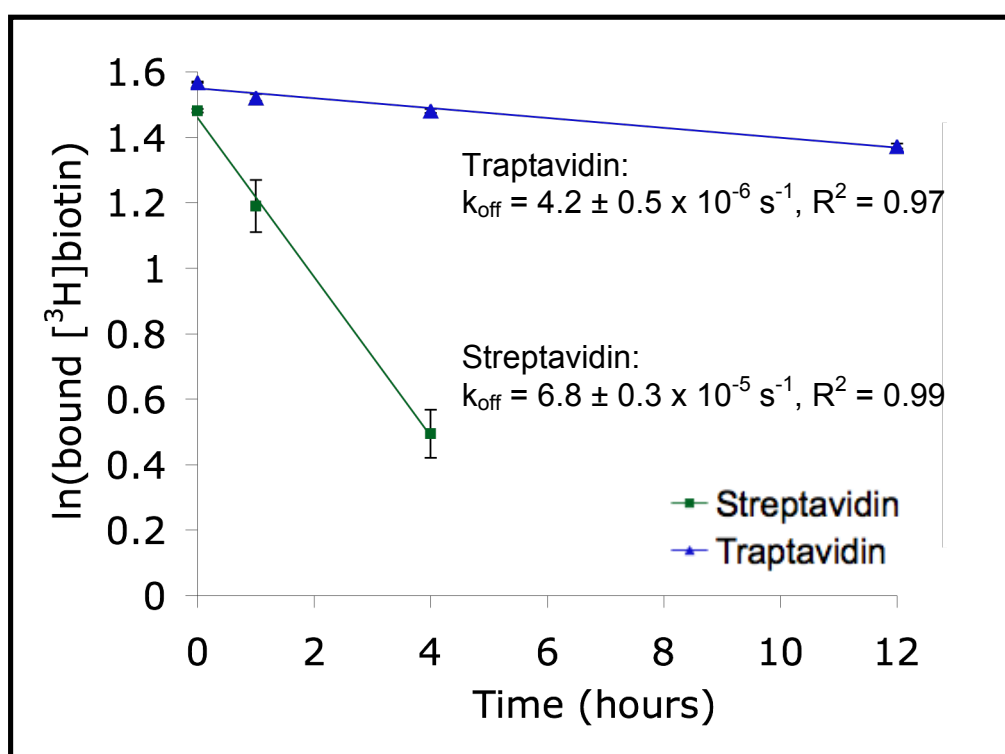


Figure 38. Tr has a greater than ten-fold slower [^3H]biotin off-rate than SA. Both proteins were bound to [^3H]biotin, followed by addition of excess, cold biotin and incubation at 37 °C. At certain timepoints, the fraction of [^3H]biotin still bound to the protein was determined. Mean of triplicate measurements ± 1 s.d.

As the B4F on-rate assays revealed all eight higher-affinity mutants, including Tr, had a lower on-rate than SA, the [^3H]biotin off-rate assay was adapted to measure the on-rate of free biotin to both Tr and SA. As with B4F, the Tr on-rate for [^3H]biotin was reduced compared to SA (Figure 39); $8.7 \pm 0.7 \times 10^6 \text{ M}^{-1}\text{s}^{-1}$ for Tr compared to $6.7 \pm 0.5 \times 10^7 \text{ M}^{-1}\text{s}^{-1}$ for SA.

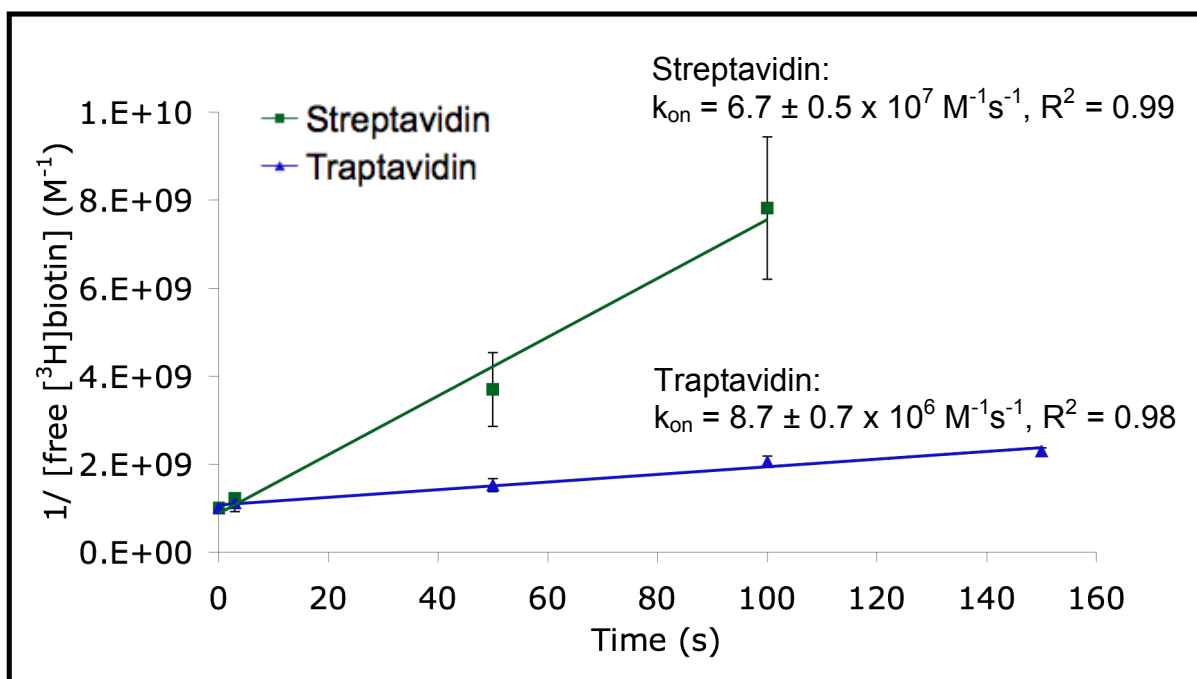


Figure 39. Tr has a slower [^3H]biotin on-rate than SA. Both proteins were bound to [^3H]biotin at time 0, and after the desired time at 37 °C, excess cold biotin was added to prevent further [^3H]biotin binding. The protein was then pulled down and the proportion of free [^3H]biotin in the solution was measured. Mean of triplicate measurements $\pm$ 1 s.d. Some error bars are too small to be visible.

From the [^3H]biotin off- and on-rates ascertained above, the equilibrium dissociation constant K_d can be calculated using $K_d = k_{\text{off}}/k_{\text{on}}$. This gives a K_d at 37 °C of $1.0 \pm 0.09 \times 10^{-12} \text{ M}$ for SA and $4.8 \pm 0.7 \times 10^{-13} \text{ M}$ for Tr.

The SA and Tr on-rates and off-rates for [^3H]biotin are summarised in Table 11.

Protein	[^3H]biotin k_{on} ($\text{M}^{-1}\text{s}^{-1}$)	[^3H]biotin k_{off} (s^{-1})	[^3H]biotin K_{d} (M)
SA	$6.7 \pm 0.5 \times 10^7$	$6.8 \pm 0.3 \times 10^{-5}$	$1.0 \pm 0.09 \times 10^{-12}$
Tr	$8.7 \pm 0.7 \times 10^6$	$4.2 \pm 0.5 \times 10^{-6}$	$4.8 \pm 0.7 \times 10^{-13}$

Table 11. Summary of [^3H]biotin assay data. All experiments were conducted at 37 °C in triplicate, mean values are shown $\pm$ 1 s.d.

4.3.3 Thermostability assays

A facile assay to assess Tr tetramer thermal stability and compare it to that of SA was conducted, involving heating the protein at the desired temperature for 3 minutes followed by SDS-PAGE (Figure 40).

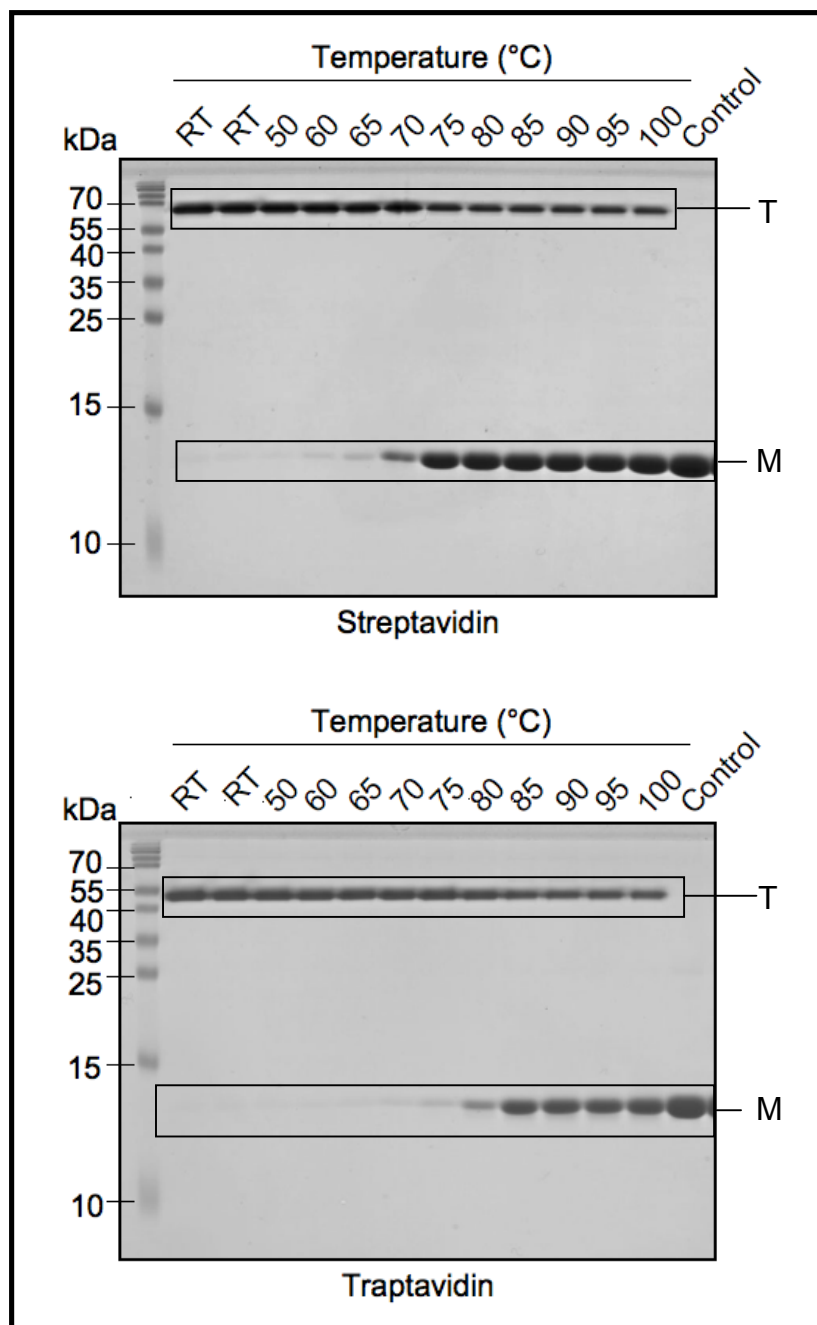


Figure 40. Visualising the tetramer-to-monomer transition of SA and Tr as a function of temperature. Each protein was incubated at the indicated temperature for 3 minutes in PBS pH 7.4 and then run on 18 % SDS-PAGE to see the distribution of tetrameric (T) and monomeric (M) protein in each sample. The positive control was heated in 6X SDS loading buffer at 95 °C for 3 minutes prior to loading.

The dissociation of tetramer to monomer as a factor of temperature was then quantified from duplicate gels and plotted (Figure 41). The mid-point of the transition from tetrameric to monomeric is ~ 10 °C higher for Tr compared to SA, indicating Tr has the higher thermostability.

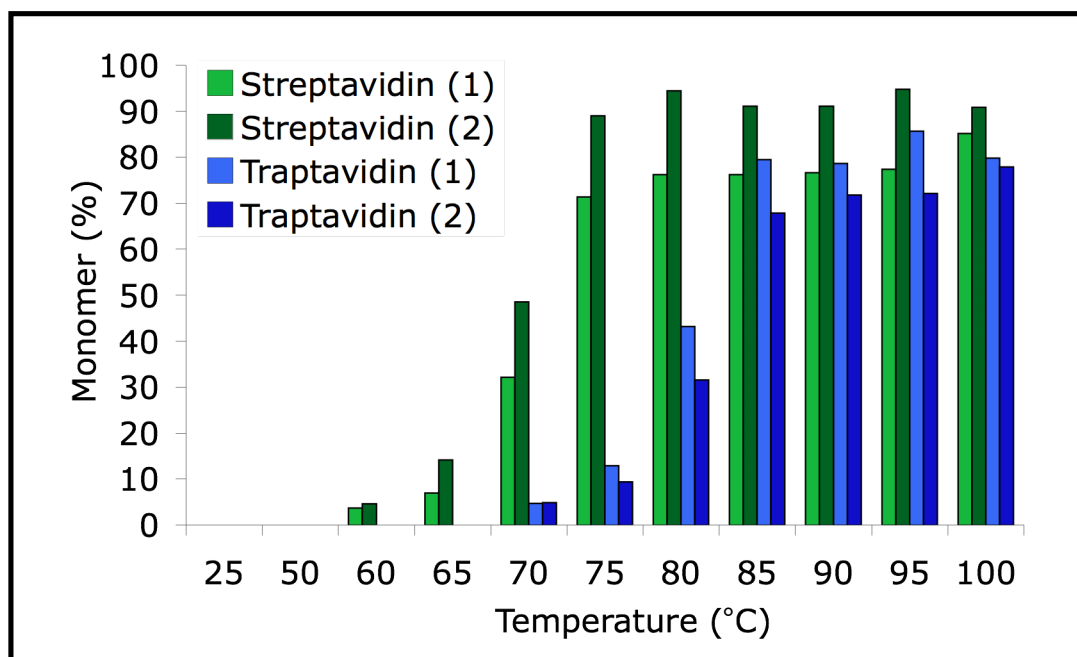


Figure 41. Traptavidin tetramer-to-monomer transition temperature is ~ 10 °C higher than streptavidin. Bands corresponding to tetrameric and monomeric protein were quantified from duplicate gels (indicated as 1 and 2) and plotted as percentage of monomeric protein.

As well as assessing the thermal stability of the Tr protein, the thermal stability of biotin-conjugate binding was also assessed, for both SA and Tr. Biotinylated DNA was bound to SA and Tr, and the complex heated at the desired temperature for 3 minutes, before running the sample on agarose gel electrophoresis to reveal the distribution of protein-bound and unbound DNA (Figure 42). At room temperature, over 90 % of the biotinylated DNA was bound to both SA and Tr, but after just 3 minutes at 50 °C, almost half of the DNA had dissociated from SA, whilst the majority was still bound to Tr. After

3 minutes at 70 °C, almost all the biotinylated DNA (99 %) was unbound and free from SA, whilst over 80 % was still bound to Tr, clearly reflecting the increased binding stability of Tr over SA.

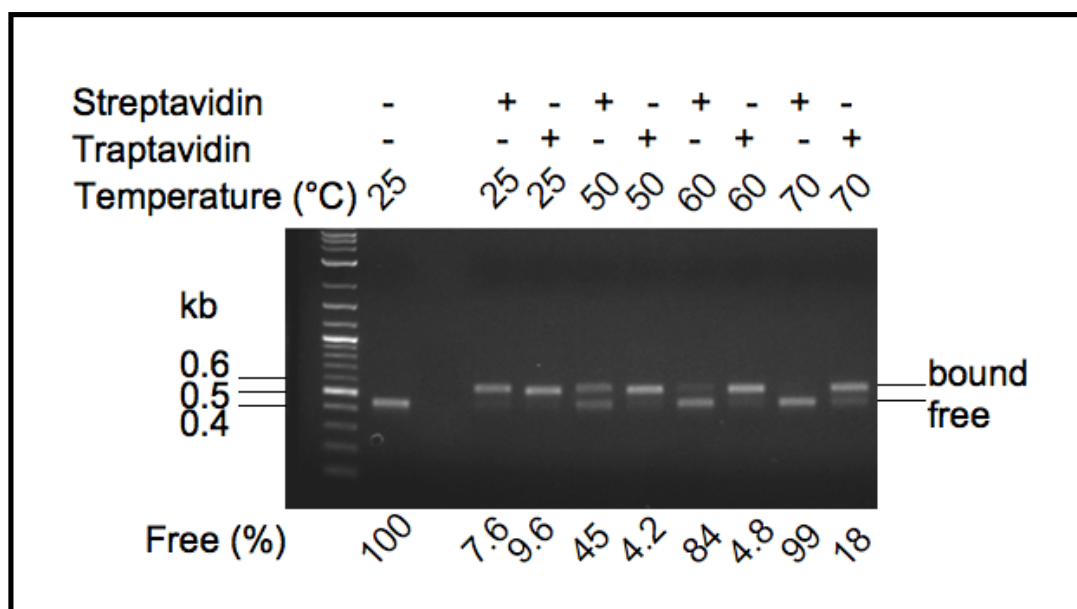


Figure 42. Tr has markedly higher thermostability of biotin-conjugate binding. SA and Tr in PBS were incubated with biotinylated DNA for 30 min at 25 °C to allow binding. Samples were then transferred into buffer containing 100 μ M biotin, 20 mM Tris acetate, 1 mM DTT, 2 mM magnesium acetate and 20 mM potassium glutamate pH 7.5 and heated for 3 minutes at the temperatures indicated followed by agarose gel electrophoresis and fluorescence imaging of DNA. The left lane is a negative control with no protein added. The percentage of free biotinylated DNA, with no protein bound, is labelled under each lane.

4.3.4 Live cell imaging

Tr contains the double mutation S52G R53D. These mutations lower the pI of the protein, from 6.09 for SA, to 5.14 for Tr (values predicted by the ProtParam program on the ExPASy proteomics server from the Swiss Institute of Bioinformatics) (Gasteiger et al., 2005). The altered charge of Tr could have affected its non-specific binding properties on cells. For this reason, the specificity of Tr on mammalian cells was compared to that of SA. This was done by fusing the acceptor peptide (AP) to the type 1 insulin-like growth factor receptor (IGF1R) and then biotinylating the AP tag with co-expressed biotin ligase (BirA-ER). The biotinylated cell-surface AP-IGF1R was then visualised by incubation with fluorescent Alexa Fluor 555-labelled SA and Tr (Figure 43). Tr gave strong staining of biotinylated IGF1R in these cells, with a staining pattern comparable to SA after a 15 minute staining incubation. However, shorter staining incubations gave a stronger signal with SA, due to Tr's slower on-rate (data not shown). Pre-blocking Tr with biotin before incubating with the cells showed that Tr exhibits minimal non-specific binding.

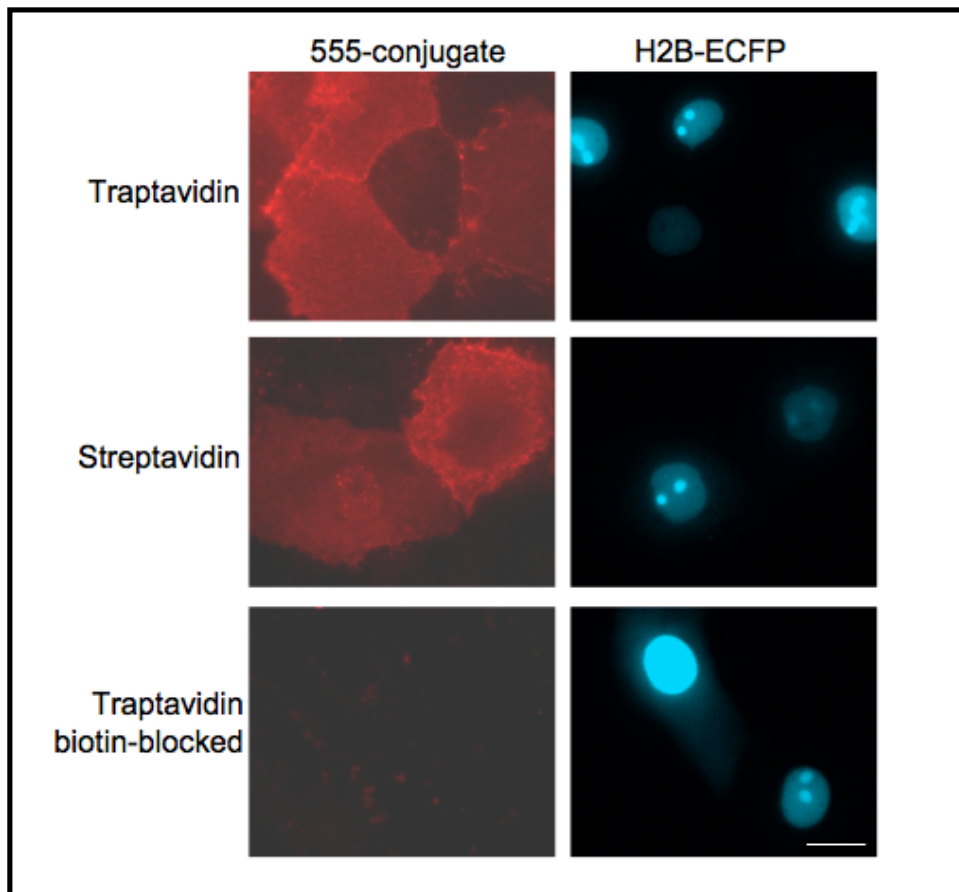


Figure 43. Traptavidin displays similar specificity as streptavidin on mammalian cells, with minimal non-specific binding. COS7 cells expressing biotinylated AP-IGF1R were incubated with Alexa Fluor 555-labelled Tr (top panel) or SA (middle panel). Alexa Fluor (red, left column) and ECFP cotransfection marker images (cyan, right column) are shown. The bottom panel is a negative control where traptavidin was pre-blocked with free biotin, showing minimal non-specific binding. Scale-bar 20 μm . Data collected by M. Howarth.

4.3.5 FtsK displacement assays

One application of the SA-biotin interaction is in the study of molecular motor translocation. In this study, we used SA and Tr to study the translocation of the molecular motor FtsK, and in the process, gained further insight into the strength of the Tr-biotin interaction in comparison to the SA-biotin interaction. In the first of the FtsK displacement assays utilising *P. aeruginosa* FtsK, PAK4, 0.5 μM PAK4 was able to displace the majority of bound SA from biotinylated DNA within 3 minutes, whilst Tr was resistant to displacement (Figure 44).

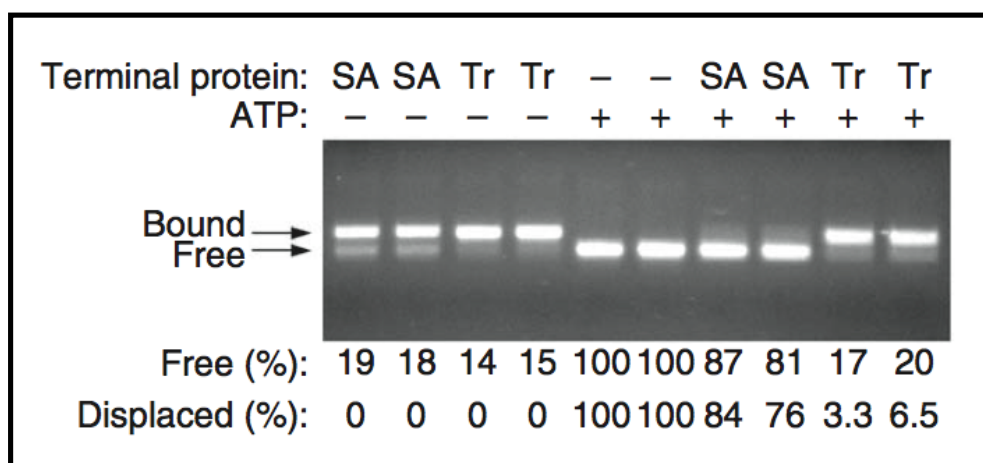


Figure 44. 0.5 μM *P. aeruginosa* FtsK PAK4 displaces the majority of streptavidin in 3 minutes, whilst most of traptavidin remains bound. The mobility shift of biotinylated DNA shown by gel electrophoresis reveals whether SA or Tr proteins have remained bound after PAK4 translocation. Controls are shown without SA/Tr protein, and without ATP (preventing PAK4 activity). The percentage free DNA and percentage of displaced protein for duplicate assays are displayed under the gel.

The apparent resistance of Tr to displacement was intriguing, and was investigated further, using a time course (Figure 45) and PAK4 concentration gradient (Figure 46). SA displacement from biotinylated DNA occurred almost immediately after PAK4 translocation began, with 31 % of the protein displaced after just 2 s, whilst over 90 % of Tr remained bound to the DNA after 300 s of PAK4 translocation (Figure 45). Increasing the concentration of PAK4 enabled greater displacement of Tr from the DNA, but Tr remained more resistant to displacement than SA, with 2 μ M PAK4 displacing 78 % of SA but only 31 % of bound Tr (Figure 46).

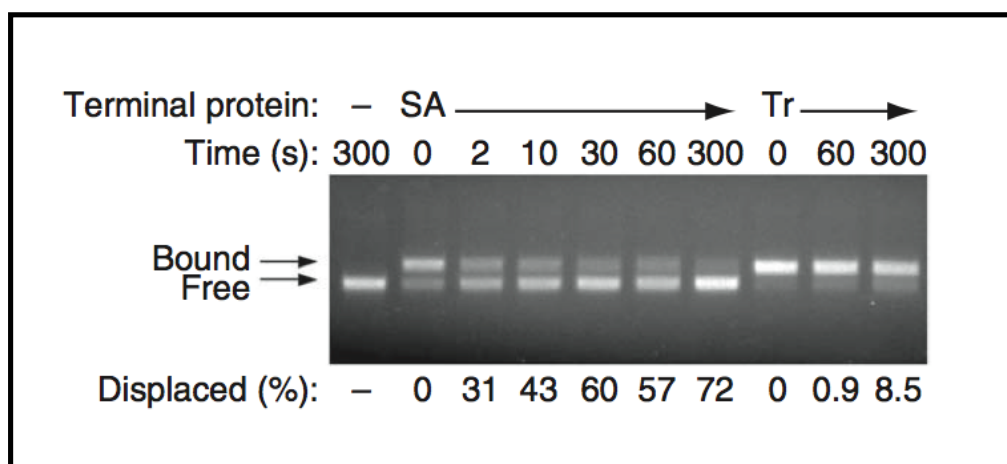


Figure 45. Time course of PAK4 displacement assay clearly shows traptavidin's increased resistance to displacement. ATP was added for the times indicated, giving control of PAK4 translocation time. Control lane without SA/Tr protein shows 100 % free DNA as FtsK does not remain bound to DNA on electrophoresis.

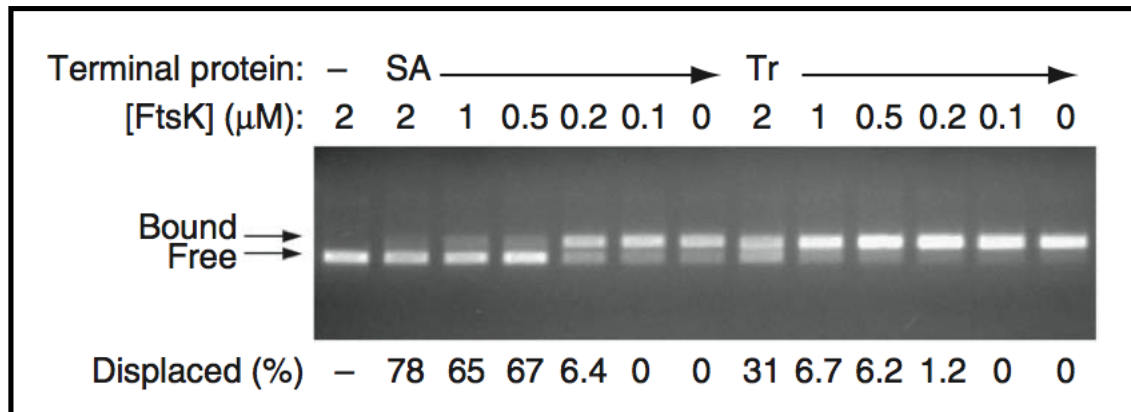


Figure 46. Increasing the concentration of FtsK PAK4 gives greater displacement of traptavidin, but traptavidin still has greater resistance to PAK4 translocation than streptavidin. The indicated concentrations of PAK4 were incubated with biotinylated DNA for 3 minutes of translocation before agarose gel electrophoresis.

The experimental set-up was also varied to test another FtsK motor, FtsK_{50C}, from *E. coli*. This assay used a DNA fragment with a terminal biotin molecule (as opposed to the internally biotinylated DNA fragment used for the PAK4 assays above). As seen with the PAK4 assays, Tr is more resistant to FtsK_{50C} displacement than SA (Figure 47).

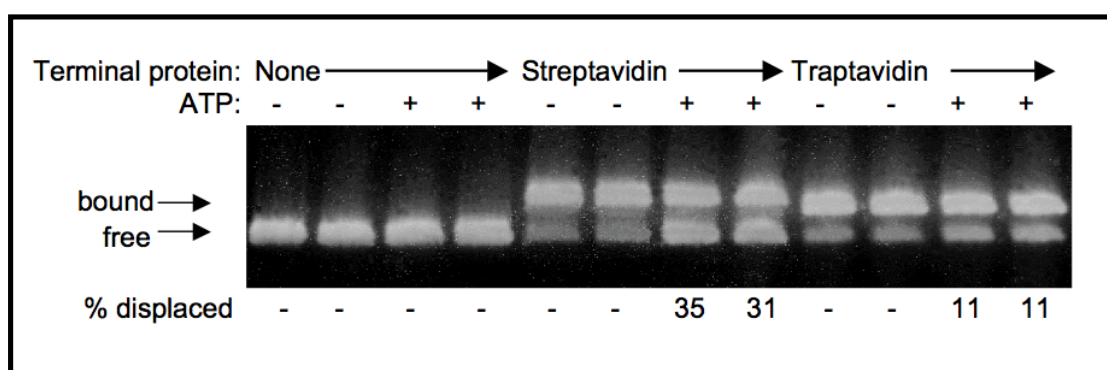


Figure 47. Traptavidin had increased resistance to *E. coli* FtsK displacement. The mobility shift of terminally biotinylated DNA was fluorescently visualised by gel electrophoresis, after 2 minutes of translocation with 1 μ M of FtsK_{50C}. The percentage of displaced protein for duplicate assays are indicated under each lane. Data collected by E. Crozat.

4.3.6 Atomic force microscopy

AFM measurements were carried out by our collaborators Calvin Chu and Vincent Moy at the University of Miami, to assess the mechanical stability of Tr at a single molecule level. A cantilever tip coated with either SA or Tr was retracted from biotinylated beads and the unbinding force measured over a range of loading rates, revealing that Tr had greater mechanical binding stability than SA over the range of loading rates ($P < 0.0001$) (Figure 48). The loading rate was restricted to give a linear range of rupture forces, indicating a single binding regime.

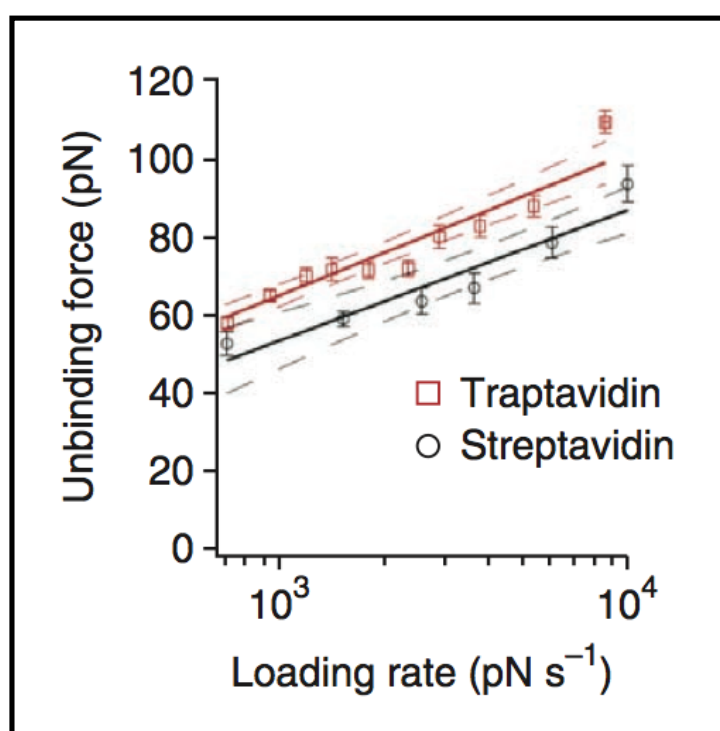


Figure 48. Traptavidin has increased mechanical stability compared to streptavidin. Single-molecule rupture forces were plotted for a range of loading rates. Means are shown ± 1 s.e.m. (SA, $n = 400$; Tr, $n = 562$), with best fits and 95 % confidence limits indicated by solid and dashed lines, respectively. Measurements were conducted at 25 °C.

Restricting the loading rates to a narrower range ($0.5 - 1.5 \times 10^3$ pN/s), the distribution of unbinding forces can be plotted (Figure 49). At a loading rate of 1.5×10^3 pN/s, the mean rupture force of the Tr-biotin interaction is 70 ± 1.1 pN, almost 10 pN higher than that of the SA-biotin interaction (61.3 ± 1.6 pN).

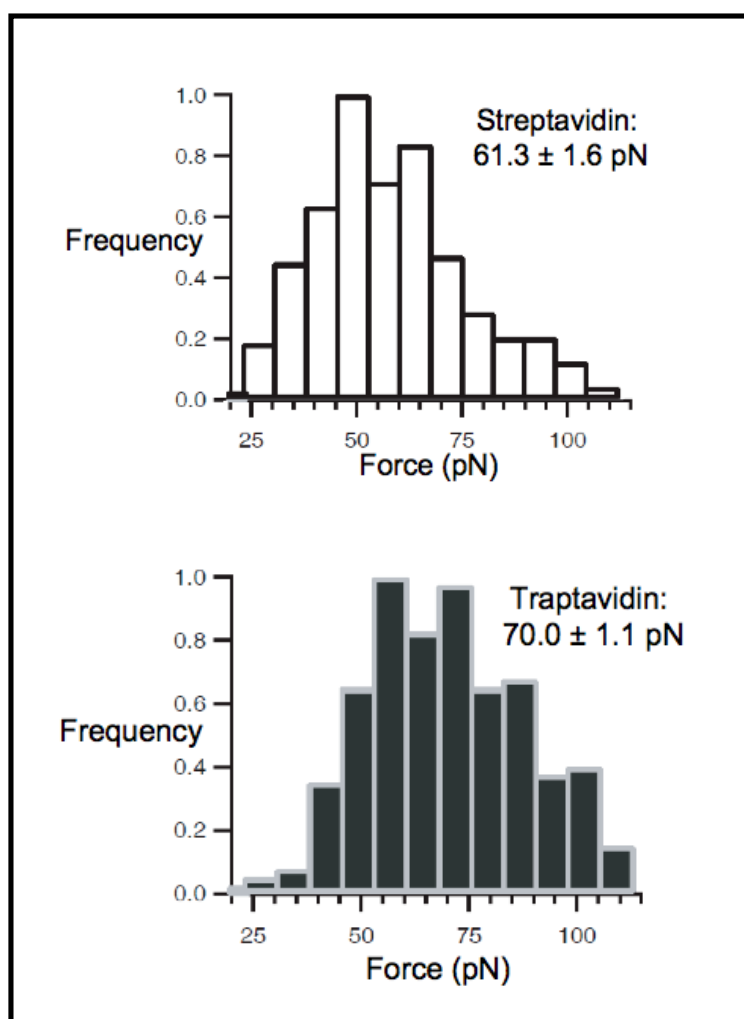


Figure 49. A distribution of biotin-binding strengths is observed by AFM, due to the influence of thermal fluctuations on traversing the activation barrier. A loading rate of 1.5×10^3 pN/s was used to investigate the SA-biotin interaction (top) and Tr-biotin interaction (bottom). Means are shown ± 1 s.e.m. (SA, $n = 265$; Tr = 249).

From the fit of loading rate to unbinding force shown in Figure 48 for SA and Tr, the relationship between the loading rate and the dissociation rate for the two proteins could be estimated (Figure 50). A clear difference emerges in the dissociation rate of the two proteins as force increases, and in particular, there is a greater than 2-fold difference in dissociation rate at forces greater than 60 pN.

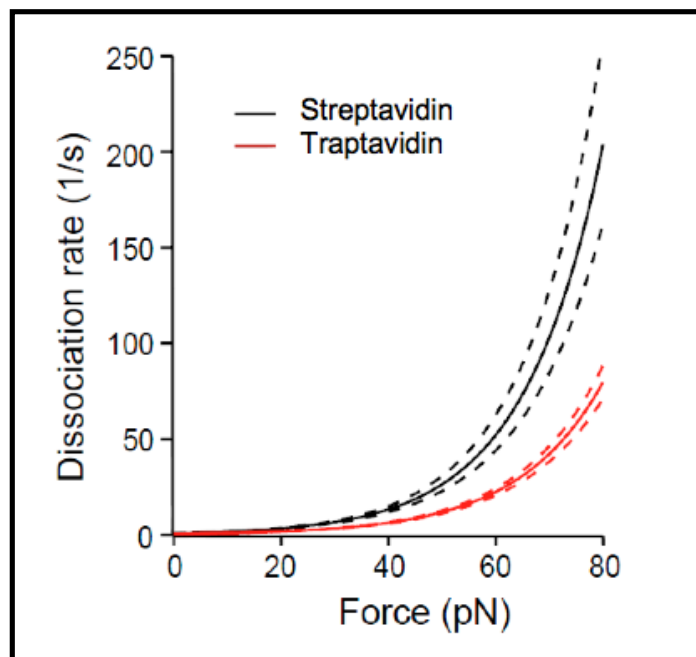


Figure 50. From the fit of loading rate to unbinding force, a prediction of the relationship between dissociation rate and applied force can be made for streptavidin and traptavidin. Dashed lines indicate values ± 1 s.e.m.

4.4 Discussion

The detailed characterisation of Tr in this chapter has confirmed the production of a SA mutant with more stable binding to biotin and biotin-conjugates.

The strong quenching of fluorescence occurring when B4F binds to SA prevents the use of the SA-B4F interaction in imaging applications, but does enable the quantification of biotin binding-sites in a sample, and also forms the basis of the initial screen conducted on each new mutant produced in this study. The B4F off-rate assay was initially conducted at 37 °C over a period of 12 hours. This assay revealed the significantly slower B4F dissociation from Tr when compared to both SA and Av (Figure 34). The initial ~ 2 % dissociation from Tr at time 0 could be ascribed to the presence of very small amounts of contaminating monomer or dimer, but gel filtration conducted prior to crystallography trials (see Chapter 5) revealed the protein purification method being used gave very clean preparations of Tr tetramer. It was also thought that the initial jump might be an artefact of a very small amount of DMSO being added with biotin at the start of the assay, but troubleshooting using the identical concentration of DMSO in the negative control still gave a ~ 2 % increase in initial dissociation (Figure 35) so this was also ruled out. An initial burst of more rapid biotin dissociation has previously been seen with Av, where the presence of some denatured protein was suggested as an explanation (Green, 1963). Figure 34 also nicely demonstrates why SA is the protein of choice, rather than the stronger biotin-binder Av, in applications

where biotin is conjugated to another molecule; SA is by far the better biotin-conjugate binder when compared to Av. A discussion of the reasons for SA's higher affinity for biotin-conjugates compared to Av is given in Chapter 5.

The initial B4F off-rate assay was extended to 96 hours at 37 °C, with very little B4F dissociation seen in this time period from Tr (in contrast to SA, from which B4F dissociated steadily). Being able to remain bound to a biotinylated molecule of interest for longer time periods at 37 °C would be beneficial in applications such as live cell imaging, as it would enable the protein of interest to be imaged and tracked within the cell for longer time periods. For example, fluorescent SA began dissociating from biotinylated nicotinic acetylcholine receptors being imaged after just 2 hours at 37 °C (Bruneau et al., 2005). However, this dissociation did not occur at the cell surface or in the extracellular environment, but rather only when the SA had been internalised into early and recycling endosomes, where the pH is between 6.3 and 6.5 (Casey et al., 2010). It is reasonable to expect that when following a biotinylated protein that enters the late endosomes (pH of 5.5) and lysosomes (pH of 4.7), the rate of dissociation would be even greater. Therefore, in order to investigate whether Tr might be more stably bound to biotinylated molecules at lower pH, the B4F off-rate assay was adapted to follow B4F dissociation from SA and Tr at pH 5. At this lower pH, B4F dissociation was faster from both SA and Tr, but dissociation from Tr was still significantly slower than from SA (Figure 37). These results suggest that the use of Tr would have advantages over SA use in live cell imaging experiments, where the molecule of interest being imaged undergoes trafficking within the cell

through compartments at lower pH. The preliminary live cell imaging experiments conducted here showed that Tr exhibited comparable staining patterns on mammalian cells as SA, with minimal non-specific binding, which also supports the prediction of Tr's increased utility in live cell imaging.

Tr's decreased off-rate and on-rate from B4F were not specific to this biotin-conjugate; the [^3H]biotin assays revealed that Tr also had a decreased off- and on-rate from free, unconjugated biotin (Figures 38 and 39), indicating slower but stronger biotin-binding to Tr. Tr had a slower biotin on-rate ($8.7 \pm 0.7 \times 10^6 \text{ M}^{-1}\text{s}^{-1}$) when compared to SA ($6.7 \pm 0.5 \times 10^7 \text{ M}^{-1}\text{s}^{-1}$, which is comparable to previously published on-rates (Hyre et al., 2006)). This means that longer incubations are required to reach equilibrium, but for applications where dissociation is the limiting factor, the on-rate of Tr should still be acceptable. In comparison to other interactions, the on-rates for protein-protein interactions are typically in the range 0.5 to $5 \times 10^6 \text{ M}^{-1}\text{s}^{-1}$ (Northrup and Erickson, 1992), and Tr's on-rate is still comparable to that of many monoclonal antibodies, such as antiluorescein antibodies ($1.3 - 9.8 \times 10^6 \text{ M}^{-1}\text{s}^{-1}$) (Kranz et al., 1982; Midelfort et al., 2004) and anti-dinitrophenol antibodies ($3.4 - 9.5 \times 10^6 \text{ M}^{-1}\text{s}^{-1}$) (James and Tawfik, 2003).

Tr's decreased biotin off-rate was desired but also rather unexpected, as despite decades of engineering efforts, there are no other examples of a SA mutant successfully designed to have an intentionally increased binding stability to free biotin. As discussed in Chapter 3, two groups have previously reported producing SA mutants with higher affinity towards a particular biotin

ligand, but these mutants have been 'stumbled' upon in their search for a SA variant with other altered properties, such as reduced antigenicity (Meyer et al., 2001) or altered desthiobiotin binding (Levy and Ellington, 2008). The off- and on-rates for Tr described here allow calculation of the equilibrium dissociation constant at 37 °C, which, at $4.8 \pm 0.7 \times 10^{-13}$ M, is half that of SA ($1.0 \pm 0.09 \times 10^{-12}$ M). This improvement in K_d is achieved from a sixteen-fold decrease in off-rate and a greater than seven-fold decrease in on-rate. Our k_{on} and k_{off} for SA are comparable with previous values (Hyre et al., 2006; Klumb et al., 1998). The K_d value for SA-biotin interaction calculated in Chapter 1 from reviewing the literature (5.9×10^{-14} M), as well as the often cited 4×10^{-14} M (Green, 1990), was determined at 25 °C, but it is known that the off-rate of SA dramatically increases with temperature (Hyre et al., 2000; Klumb et al., 1998).

Tr's slower off-rate from both B4F and free biotin reflects the increased kinetic stability of the bound Tr complex once formed; however, we also sought to characterise the thermal stability. Thermostability assays showed Tr had ~ 10 °C higher tetramer thermostability than SA before dissociating into monomers, as well as revealing Tr possessed much higher biotin-conjugate binding stability at high temperatures than SA. This increased thermostability of both the Tr tetramer and its biotin-conjugate binding suggests the mutant would have an advantage over SA in applications that involve extended time periods at higher temperatures, such as DNA 'sequencing by synthesis' (Braslavsky et al., 2003) and live cell imaging investigations. Sequencing by synthesis is one of the most successful and widely adopted next-generation technologies

for DNA sequencing, together with nanopore sequencing (Branton et al., 2008). DNA sequencing by synthesis involves the detection of single bases as they are incorporated into growing DNA strands, by using bases fluorescently labelled at their terminal phosphate. In order to maximise processivity of the polymerase, enabling longer reads, SA can be used as a 'clamp', to both trap the DNA being sequenced in the polymerase and to also allow orientated immobilisation onto a surface (Williams et al., 2008). However, sequencing at the optimum temperature for polymerase activity (74 °C) causes SA to dissociate from the DNA-polymerase complex, and so sequencing has to take place at lowered temperatures to prevent this dissociation, such as 54 °C, at which polymerase activity is only a third of its maximum (Williams et al., 2008). The use of Tr in this application, with its increased thermostability, would clearly be advantageous over SA, as it would enable the use of higher temperatures, giving increased polymerase activity.

The FtsK displacement assays used in this study were extremely useful, both for investigating the motor's translocation and for comparing the mechanical stability of the Tr-biotin and SA-biotin interactions. As FtsK translocates along DNA, it is capable of generating a force > 60 pN (Pease et al., 2005) and can disrupt many DNA-proteins interactions in its path, including the SA-biotin interaction (Chivers et al., 2010). The ability of Tr to resist FtsK displacement to a much greater extent than SA demonstrated the increased mechanical strength of the Tr-biotin interaction compared to SA-biotin. The observation that increasing the FtsK concentration enabled greater displacement of Tr also indicated that multiple FtsK motors could be loading onto the DNA and

cooperating to exert a stronger pushing force. Tr, SA and a mutant SA with weaker biotin-binding, A86D H87G (see Appendix A1), have also been used to study the translocation properties of FtsK mutants, to ascertain the mechanism of translocation (Croizat et al., 2010). In these studies, three FtsK_C domains were covalently linked together to form trimers, which were then allowed to dimerise to form hexamers. By introducing mutations in defined subunits, mutant FtsK hexamers were produced that had a defined number of catalytically inactive subunits (Figure 51).

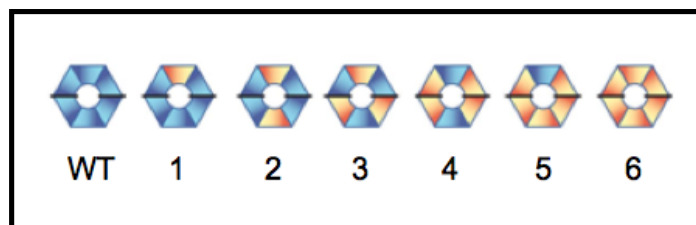


Figure 51. Mutant FtsK hexamers can be used in addition to SA and Tr to investigate the translocation method of FtsK. Mutated subunits are coloured yellow and wildtype subunits are coloured blue. The number of mutated subunits in the hexamer are labelled. Figure adapted from that which was previously published (Croizat et al., 2010).

FtsK mutants with two inactivated ATPase motors were capable of translocating at the same speed as wildtype FtsK, which ruled out a ‘stochastic firing’ model where subunits in the hexamer ‘fire’ randomly (a linear decrease in translocation speed as the number of inactivated subunits in the hexamer increased would be expected in this case) and also a ‘concerted’ model where every subunit is required to ‘fire’ for translocation to occur (an inactive subunit would prevent any translocation in this case) (Figure 52). Despite translocating at the same speed as wildtype FtsK, a mutant FtsK with two inactive subunits was unable to displace SA ‘roadblocks’ on the DNA, whilst wildtype FtsK could. This separation of translocation

velocity from force generation suggested an ‘escorted translocation’ model (where only a subset of subunits make contact with the DNA at any time) (Croizat et al., 2010).

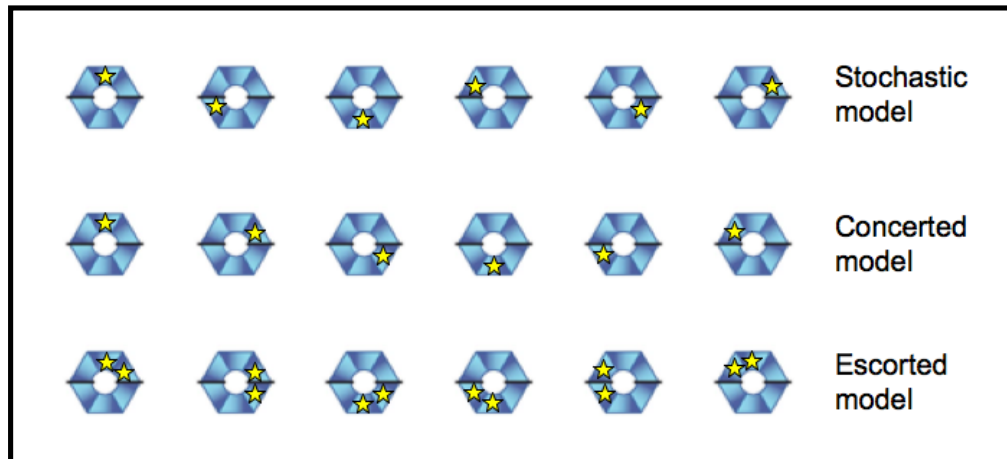


Figure 52. Various translocation models proposed for FtsK. There is random ‘firing’ (ATP hydrolysis) of the subunits in the hexamer in the stochastic model (top), whilst every subunit is required to ‘fire’ in the concerted model (middle). The escorted translocation model (bottom), where only a subset of subunits ‘fire’ at any time is currently the more favoured model (Croizat et al., 2010).

As well as helping to elucidate the translocation method of FtsK, the resistance of Tr to molecular motor displacement could also allow further investigation of the translocation method using single-molecule imaging. For example, it is not known for how long multiple FtsK motors will build-up on the DNA at the site of an immovable ‘roadblock’ on the DNA (such as Tr) before dissociating. Single-molecule imaging of fluorescently-labelled FtsK translocating along a DNA substrate with Tr bound at a defined site could be utilised, for example, with the ‘DNA curtains’ method developed by Eric Greene’s lab (Visnapuu and Greene, 2009).

Investigating the Tr-biotin interaction at a single molecule level with AFM also revealed that Tr had increased mechanical strength compared to SA. The rupture forces of the SA-biotin and Tr-biotin interactions were measured over a wide range of loading rates, and the mean rupture force for the Tr-biotin interaction was almost 10 pN higher than that for SA-biotin. There is a distribution of binding strengths in the plots due to the importance of thermal fluctuations in traversing the activation barrier to dissociation (Evans and Ritchie, 1997). Using a different force measurement method (the biomembrane force probe technique) similarly broad force-histograms were obtained for the SA-biotin interaction (Merkel et al., 1999). From the plot of dissociation rate against rupture force, a clear difference emerges in the dissociation rate of the two proteins as force increases. Specifically, there is a greater than 2-fold difference in dissociation rates at forces greater than 60 pN, making it conceivable that an increase in rupture force from 61 pN (for the SA-biotin interaction) to 70 pN (for the Tr-biotin interaction) would have a large effect on the dissociation rate.

With both an increased mechanical stability and thermal stability, the use of Tr rather than SA could prove to be advantageous in targeted radionuclide therapy (TRT). As explained in detail in Section 1.4.1, TRT utilises the SA-biotin interaction to facilitate the delivery of a radionuclide specifically to monoclonal antibody-labelled tumour cells (Boerman et al., 2003). For the radioactivity to be retained at the tumour site, the SA-biotin interaction must withstand an extended time period at 37 °C, as well as being resistant to any shear forces present, e.g. from blood flow in a capillary. Both of these

conditions have already been shown to accelerate the dissociation of biotin from SA (Klumb et al., 1998; Pierres et al., 2002). The Tr-biotin interaction would be more stable than SA-biotin under these conditions, and therefore, as long as the decreased on-rate of Tr did not pose a problem, Tr could replace the use of SA in TRT.

After completing the Tr characterisation presented in this chapter, it is clear that Tr has the potential to replace SA in applications where the experimental conditions pose more of a challenge to the SA-biotin interaction, such as at higher temperature, lower pH or under mechanical stress.

When attempting to explain the origins of Tr's properties, one could hypothesise that the altered surface electrostatic charge in Tr could have increased its stability, as it has been shown that optimising the surface charge-charge interactions of certain enzymes increases their stability without altering their activity (Gribenko et al., 2009). Also, due to the important role L3/4 has in generating the high biotin-binding affinity, and the proximity of the S52G R53D mutations to L3/4, it was hypothesised that the mutations were affecting the loop's flexibility. This would explain the decreased biotin and biotin-conjugate on- and off-rates of Tr. It was clear that having structures of both apo-Tr and biotin-bound Tr would be extremely useful in shedding more light on how the S52G R53D mutations in Tr were exerting their effect. Hence, X-ray crystallography was conducted in an attempt to explain the origins of the increased binding stability of Tr (Chapter 5).

Chapter 5: The Structures of Apo-Traptavidin and Biotin-Bound Traptavidin

5.1 Introduction

As of 30th July 2011, there was a total of 147 structures deposited in the PDB that contained SA (both wildtype and mutated), either in apo-form or ligand-bound (both biotin and non-natural ligands). Of these, 26 structures were of wildtype SA, either in apo-form or biotin-bound (detailed in Table 12). The number of crystal structures in the PDB, together with the wide range of crystallisation conditions used to obtain these structures and the resolutions achieved, suggested SA crystallises readily, and so it was hoped that Tr would also readily form high-quality crystals. Acquiring high-resolution structures of both apo-Tr and biotin-Tr would enable detailed comparisons to be made to apo-SA and biotin-SA structures, to shed light onto how the S52G R53D mutations in Tr exert their effect on biotin-binding stability and protein thermostability (see also Chapters 3 and 4).

As discussed in Chapter 1, the key structural features of SA and its complex with biotin include: the tetrameric quaternary structure of SA; the deep biotin-binding pocket in SA that is highly complementary to the shape of biotin; the numerous van der Waals interactions that form between biotin and SA, including interactions between biotin and the four conserved tryptophans in the binding pocket; the extensive hydrogen-bonding network that forms between biotin and eight conserved residues in the binding pocket; and a surface loop connecting β -strands 3 and 4 (L3/4) which is stabilised in a

'closed' conformation on biotin-binding, acting as a lid over the binding pocket.

All of these features are seen in the structures listed in Table 12.

Description	PDB code	Resolution (Å)	pH	Reference
Apo-SA structures				
T7-tagged full-length SA	2BC3	1.54	4.5	Le Trong et al., 2006
Apo-SA	2RTA	1.39	2.97	Katz, 1997
Apo-SA	2RTB	1.50	3.32	Katz, 1997
Apo-SA	2RTC	1.50	3.60	Katz, 1997
Core apo-SA	2IZA	1.46	2.00	Katz, 1997
Core apo-SA	2IZB	1.20	3.12	Katz, 1997
Core apo-SA	2IZC	1.40	2.00	Katz, 1997
Core apo-SA	2IZD	1.60	3.00	Katz, 1997
Core apo-SA	2IZE	1.57	3.08	Katz, 1997
Core apo-SA	1SWA	1.90	4.5	Freitag et al., 1997
Core apo-SA	1SWB	1.85	7.5	Freitag et al., 1997
Core apo-SA	1SWC	1.80	4.5	Freitag et al., 1997
Core apo-SA	1SLF	1.76	5.6	Katz, 1995
Biotin-SA structures				
Core SA with biotin	1MK5	1.40	4.5	Hyre et al., 2006
Core SA with biotin	2RTD	1.65	1.39	Katz, 1997
Core SA with biotin	2RTE	1.50	1.90	Katz, 1997
Core SA with biotin	2RTF	1.47	2.00	Katz, 1997
Core SA with biotin	2RTG	1.39	2.40	Katz, 1997
Core SA with biotin	2IZF	1.58	4.0	Katz, 1997
Core SA with biotin	2IZG	1.36	2.0	Katz, 1997
Core SA with biotin	2IZH	1.36	10.44	Katz, 1997
Core SA with biotin	2IZI	1.50	2.53	Katz, 1997
Core SA with biotin	2IZJ	1.30	3.50	Katz, 1997
Core SA with two biotins	1SWD	1.90	4.5	Freitag et al., 1997
Core SA with four biotins	1SWE	2.06	4.5	Freitag et al., 1997
Core SA with biotin	1STP	2.60	7.8	Weber et al., 1989

Table 12. Summary of all wildtype SA structures, either in apo-form or biotin-bound, deposited in the PDB. Correct as of 30th July 2011.

Due to the close proximity of the S52G R53D mutations to L3/4 (residues 45-52), and the well-documented role of L3/4 in generating high-affinity binding, it was hypothesised that the mutations may affect the flexibility of L3/4, and so particular attention was paid to the conformation of this region in the published SA structures in Table 12. In a series of SA structures solved at extremes of

pH by B. Katz (Katz, 1997), L3/4 was found stabilised in a 'closed' conformation in apo-SA, and then remained in this 'closed' conformation in the biotin-bound protein. However, it is widely commented in the literature that L3/4 is disordered in an 'open' conformation in the apo-forms of SA, and is stabilised upon biotin-binding into a 'closed' conformation (Freitag et al., 1997). It is indeed the case that in the SA structures solved in the more physiologically relevant pH range (pH 4.5 to 7.5), L3/4 is found disordered in an 'open' conformation in the apo-state and found stabilised and 'closed' in the biotin-bound state.

After solving the crystal structures of both apo-Tr and biotin-Tr, I undertook a detailed analysis of the structures, including structural comparisons using backbone alignments, measurements of bond lengths and angles in bound biotin and measurements of the hydrogen bond lengths to biotin. Flexibility analysis was based on *B* factors, which are atomic displacement parameters that reflect the thermal motion and disorder of each atom in the crystal structure (Yuan et al., 2005b). An area of a protein with a low *B* factor corresponds to a well-ordered region, whilst an area with a high *B* factor corresponds to a flexible region. However, *B* factor analysis must take place alongside observing the electron density for a structure, as variations in occupancies will result in high *B* factors. A chimeric Tr tetramer, Tr1D3, was also produced during this investigation, consisting of one Tr subunit and three 'dead' SA subunits that cannot bind biotin. The characterisation of this monovalent Tr revealed information on the role of the intersubunit contacts in

Tr, as well as providing evidence that Tr1D3 could be as useful a biotechnological tool as tetravalent Tr.

5.2 Materials and methods

5.2.1 Expression and purification of proteins used in this study

Tetravalent wildtype streptavidin (SA), tetravalent 'dead' streptavidin (D4), tetravalent traptavidin (Tr) and monovalent traptavidin (Tr1D3) were all used in this study.

5.2.1.1 Tetravalent traptavidin expression and purification

His₆-tagged Tr was expressed in *E. coli* BL21 (DE3) RIPL cells and purified from inclusion bodies as described in Chapter 3. Prior to crystallisation tray set-up, size-exclusion chromatography was carried out as a final purification step and to assess the purity of the purification protocol. The Tr protein sample was concentrated in a 10 kDa MWCO Vivaspin 6 concentrator (Generon Ltd., UK) prior to loading onto a XK 26/100 column packed with Sephadex[™] G-75 (GE Healthcare, UK), equilibrated in 50 mM Tris-HCl, 0.5 M NaCl, 10 mM EDTA pH 7.5. Using the chromatogram as a guide, fractions containing tetrameric Tr were pooled and then concentrated in a 10 kDa MWCO Vivaspin 6 concentrator to a final concentration of 19 mg/mL in 50 mM Tris-HCl, 0.5 M NaCl, 10 mM EDTA pH 7.5 for crystallisation.

Biotin-bound Tr was obtained by incubation with a 4-fold molar excess (compared to the number of binding sites) of biotin (Acros Organics, UK) at 4 °C overnight prior to crystal tray set-up.

5.2.1.2 Monovalent traptavidin expression and purification

Monovalent Tr (Tr1D3) was produced by combining dead (D) and Tr subunits in a 3:1 molar ratio in guanidine hydrochloride, followed by refolding by rapid dilution into PBS in a 'mixed-refold' method as described for monovalent streptavidin elsewhere (Howarth et al., 2006; Howarth and Ting, 2008). However, the order of SDS-PAGE mobility was reversed for Tr compared to SA (mobility of Tr2D2 >Tr1D3>D4).

5.2.1.3 Tetraivalent 'dead' streptavidin expression and purification

Tetraivalent 'dead' streptavidin (D4) was generated from the inclusion body purification of the triple SA mutant N23A S27D S45A, which has negligible biotin binding (Howarth et al., 2006).

D4 protein was expressed in *E. coli* BL21 (DE3) RIPL cells as described in Chapter 3 for the higher-stability SA mutants. After protein induction was confirmed by SDS-PAGE, inclusion bodies were isolated from a 750 mL culture cell pellet by rocking with 10 mL 300 mM NaCl, 50 mM Tris, 5 mM EDTA, 0.8 mg/mL lysozyme, 1 % Triton X-100 pH 7.8 for 30 min at 25 °C followed by 9 min pulsed sonication on ice at 40 % amplitude on a Sonics Vibra-Cell sonicator. Following centrifugation at 27,000 g for 15 min at 4 °C, the inclusion body pellet was washed three times in 10 mL 100 mM NaCl, 50 mM Tris, 0.5 % Triton X-100 pH 7.8 and then solubilised in 4 mL of 6 M guanidine hydrochloride (GuHCl) pH 1.55. Denatured protein in GuHCl was then centrifuged at 13,000 g for 15 min at 4 °C, and the clear supernatant refolded by rapid dilution into forcefully stirring PBS (137 mM NaCl, 2.7 mM

KCl, 8.1 mM $\text{Na}_2\text{HPO}_4 \cdot 2\text{H}_2\text{O}$, 1.8 mM KH_2PO_4 pH 7.4) at 4 °C, followed by stirring overnight. Refolded D4 protein was then centrifuged at 17,700 g for 15 min at 4 °C to precipitate any misfolded proteins.

Soluble refolded D4 was then precipitated out of solution using ammonium sulfate. 62.7 g of ammonium sulfate was added slowly (~ 10 g at a time) to the D4 solution and left stirring at 4 °C for 3 hours. The solution was then filtered through tissue paper and a further 59 g of ammonium sulfate added to the flow-through and left stirring overnight at 4 °C. Following centrifugation at 17,700 g at 4 °C for 15 min, the precipitate (D4) was resuspended in a minimal volume of PBS (~ 3 mL) and then dialysed three times against PBS.

5.2.2 Crystallisation

5.2.2.1 Initial screening of crystallisation conditions

Apo-Tr and biotin-Tr samples were subjected to initial crystallisation trials using commercially available screens (the sparse matrix 'Structure Screens I and II' from Molecular Dimensions, UK and the Index™ screen from Hampton Research, USA). Initial screening was performed at 18 °C using the sitting-drop vapour-diffusion method. Screening plates were prepared by dispensing 40 µL of each screen condition as the reservoir solution into 3-well SwissSci crystallisation plates (Hampton Research, USA) using a Tecan Genesis 150 Robotic Workstation (Tecan, UK). 66 nL, 100 nL and 133 nL drops of protein samples were set up using a mosquito® nanolitre pipettor (TTP LabTech, UK). Plates were incubated at 18 °C and checked manually for crystal formation by microscopy using a Leica MZ16 stereomicroscope (Meyer Instruments, USA) after one, three and seven days.

5.2.2.2 Optimisation of crystallisation conditions

Optimisation was conducted by screening around the conditions in the initial screens that yielded crystals by varying the salt and precipitant concentrations. These custom optimisation screens were designed using in-house 'Traymaker' software written by M. Noble. Optimisation plates were set-up as above, dispensing 40 µL of reservoir solution and 500 nL of the protein sample using the Tecan and mosquito® robots respectively.

Manual inspection of the optimisation screens revealed crystals in almost all of the conditions, appearing after two days. Plates were kept at 18 °C and

transported to the beamline, where individual crystals were mounted into a 0.2 μm loop, dipped for a few seconds in an appropriate cryoprotectant solution immediately followed by immersion in liquid nitrogen, and then mounted on the goniometer head. The selection of a crystal for exposure and choice of cryoprotectant used was made via a 'trial and error' method, based on the quality of the diffraction pattern observed from a short exposure in the beamline.

5.2.2.3 Final crystallisation conditions used: apo-Tr

Crystals were obtained at 291 K by the sitting-drop vapour-diffusion method. Apo-Tr crystals belonged to the space group $I-4_1$ ($a=b=57.59$ Å, $c=183.35$ Å), with two Tr monomers present in the asymmetric unit. Apo-Tr crystals were obtained from a 4 μL drop of 19 mg/mL apo-Tr solution and a reservoir solution of 12 % (v/v) poly(ethylene glycol) (PEG) 8000, 9 % (v/v) ethylene glycol and 0.1 M HEPES pH 7.5, with crystals appearing after two days and reaching optimum size after four days. Prior to data collection, crystals were briefly soaked in a cryoprotectant solution of 12 % (v/v) PEG 8000, 30 % (v/v) ethylene glycol and 0.1 M HEPES pH 7.5 before immersion into liquid nitrogen.

5.2.2.4 Final crystallisation conditions used: biotin-Tr

Biotin-bound crystals belonged to the space group $P4_22_12$ ($a=b=57.34$ Å, $c=77.55$ Å), with one monomer present in the asymmetric unit. Biotin-Tr crystals were obtained from a 4 μL drop of 19 mg/mL biotin-Tr solution and a reservoir solution of 27 % (v/v) PEG 4000, 0.25 M MgCl_2 and 0.1 M Tris-HCl

pH 8.5. Crystals appeared after two days and reached optimum size after four days. Prior to data collection, crystals were briefly soaked in a cryoprotectant solution of 27 % (v/v) PEG 4000, 0.25 M MgCl_2 and 25 % glycerol and 0.1 M Tris-HCl pH 8.5 before immersion into liquid nitrogen.

5.2.2.5 Diffraction data collection

Crystallographic data were collected using an Oxford Cryosystems 700 series Cryostream on an ADSC Quantum 315 CCD (charge-coupled device) detector at 100 K with an oscillation range of 0.5° at beamline IO2 at the Diamond Light Source, Harwell, U.K.

5.2.2.6 Structure solution and refinement: apo-Tr

Data were indexed and integrated using MOSFLM, and then scaled and merged using SCALA from the CCP4 program suite (1994). The structure was phased by molecular replacement with the program Phaser (McCoy et al., 2007), using wild-type core SA (PDB code 1SWB) (Freitag et al., 1997) as the search model. Two monomers were present in the asymmetric unit. 5 % of the reflection data were set aside for the calculation of R_{free} using the Freerflag program in CCP4 (Brunger, 1992). The model was built in Coot (Emsley and Cowtan, 2004) and refined in PHENIX refine (Adams et al., 2002a), with incorporation of the twinning operator $(-h, k, -l)$ with a twinning fraction of 49.7 %. Throughout the refinement, all results were included from 27.47 Å resolution to the highest limit (1.45 Å) and anisotropic temperature factors were refined. The model was evaluated with MolProbity (Chen et al., 2010), which gave Ramachandran statistics with 98 % of residues in favoured

regions and no outliers. The diffraction-data precision indicator (DPI) indicated that the agreement between the model and the X-ray data for apo-Tr was 0.056 Å. The final apo-Tr model was deposited in the PDB using the accession code 2Y3E.

5.2.2.7 Structure solution and refinement: biotin-Tr

The results were indexed and integrated using MOSFLM, and scaled and merged using SCALA from the CCP4 program suite. Intensities were converted into structure factors using the truncate program (French and Wilson, 1978) in CCP4 and no signs of twinning were observed in the results. The structure was phased by molecular replacement using Phaser (McCoy et al., 2007), with the apo-Tr structure used as the search model. One monomer was present in the asymmetric unit. 5 % of the reflection data were set aside for the calculation of R_{free} using the Freerflag program (Brunger, 1992) in CCP4. Model building was performed in Coot (Emsley and Cowtan, 2004) and refinement, including rounds of simulated annealing refinement, in PHENIX refine (Adams et al., 2002a). Throughout the refinement, all data were included from 38.78 Å resolution to the highest limit (1.49 Å) and anisotropic temperature factors were refined. The presence of a well-defined biotin model in the binding pocket was clearly visible in the initial $F_{\text{obs}} - F_{\text{calc}}$ electron density maps. Biotin co-ordinates from wild-type core SA with biotin at 1.4 Å (PDB code 1MK5) (Hyre et al., 2006) were used as a template for the refinement and the weighting of co-ordinate refinement was optimized by PHENIX refine as the refinement progressed. The model was evaluated with MolProbity (Chen et al., 2010), which gave Ramachandran statistics with 98 % of

residues in favoured regions and no outliers. The DPI indicated that the agreement between the model and the X-ray data for biotin–Tr was 0.051 Å. The final biotin-Tr model was deposited in the PDB using the accession code 2Y3F.

To assess whether model bias was introduced into the biotin–Tr model as a result of using the apo-Tr structure as a search model for molecular replacement, a simulated annealing composite omit map was calculated using PHENIX AutoBuild (Terwilliger et al., 2008), with the starting phases provided by the apo-Tr structure.

5.2.3 Structure analysis

Structural overlays and colouring according to *B* factor were done in PyMOL (Schrodinger, 2010). Backbone alignments were calculated from main-chain atoms of apo-Tr chain A, biotin–Tr, apo-core-SA at pH 4.5 (PDB code 1SWA and 1SWC), apo-core SA at pH 7.5 (PDB code 1SWB) and biotin–core-SA at pH 4.5 (PDB codes 1SWD and 1SWE) (Freitag et al., 1997). Plots of mean *B* factors of main-chain atoms against residue number were produced using the BAVERAGE program in CCP4. Measurements of biotin geometry and hydrogen bond lengths were carried out in Coot (Emsley and Cowtan, 2004) using the biotin-Tr model (PDB code 2Y3F) and the biotin-SA models 1SWD, 1SWE (Freitag et al., 1997) and 1MK5 (Hyre et al., 2006). The DPI was assessed using the program SFCHECK from CCP4.

5.2.4 Biotin-4-fluorescein off-rate assay

The off-rate of B4F from SA, Tr and Tr1D3 was measured at 37 °C and pH 7.4 as described in Section 3.2.3.

5.2.5 Thermostability analysis of proteins

The thermostability of Tr, D4 and T1D3 in PBS was ascertained as described in Section 3.2.5.

5.3 Results

5.3.1 Traptavidin protein purification

Prior to crystallisation, size-exclusion chromatography was used as an additional final step to the purification protocol normally used for Tr production. The chromatogram from this size-exclusion chromatography indicated that the standard Tr purification protocol used up until this point did produce a clean sample of tetrameric Tr, with no contaminating monomer or dimer present (results not shown).

After size-exclusion chromatography, Tr was concentrated to 19 mg/mL and no precipitate was seen whilst concentrating.

5.3.2 Crystallisation

5.3.2.1 Initial screening and optimisation of crystallisation conditions

Initial crystallisation trials using commercially available sparse matrix screens were very encouraging, as crystals were visible after three days under a great number of conditions. The conditions that gave the largest and most dispersed crystals are listed in Table 13.

Protein	Conditions giving good crystal formation in initial screening
Apo-Tr	A1: 2 M Na formate, 0.1 M Na acetate pH 4.6 A2: 18 % PEG 8000, 0.2 M Ca acetate, 0.1 M Na cacodylate pH 6.5 A3: 10 % PEG 8000, 0.1 M HEPES, 8 % ethylene glycol pH 7.5 A4: 0.1 M NaCl, 0.1 M Bis-Tris, 1.5 M ammonium sulfate pH 6.5
Biotin-Tr	B1: 30 % PEG 4000, 0.1 M Tris-HCl, 0.2 M MgCl ₂ pH 8.5

Table 13. Initial crystallisation trials using commercially available screens revealed that many conditions gave well-formed crystals.

Finer screening around the conditions listed in Table 13 was then conducted, using protein drops of a larger volume and varying the salt and precipitant concentrations. The optimisation screens of the initial crystallisation conditions are shown in Figure 53.

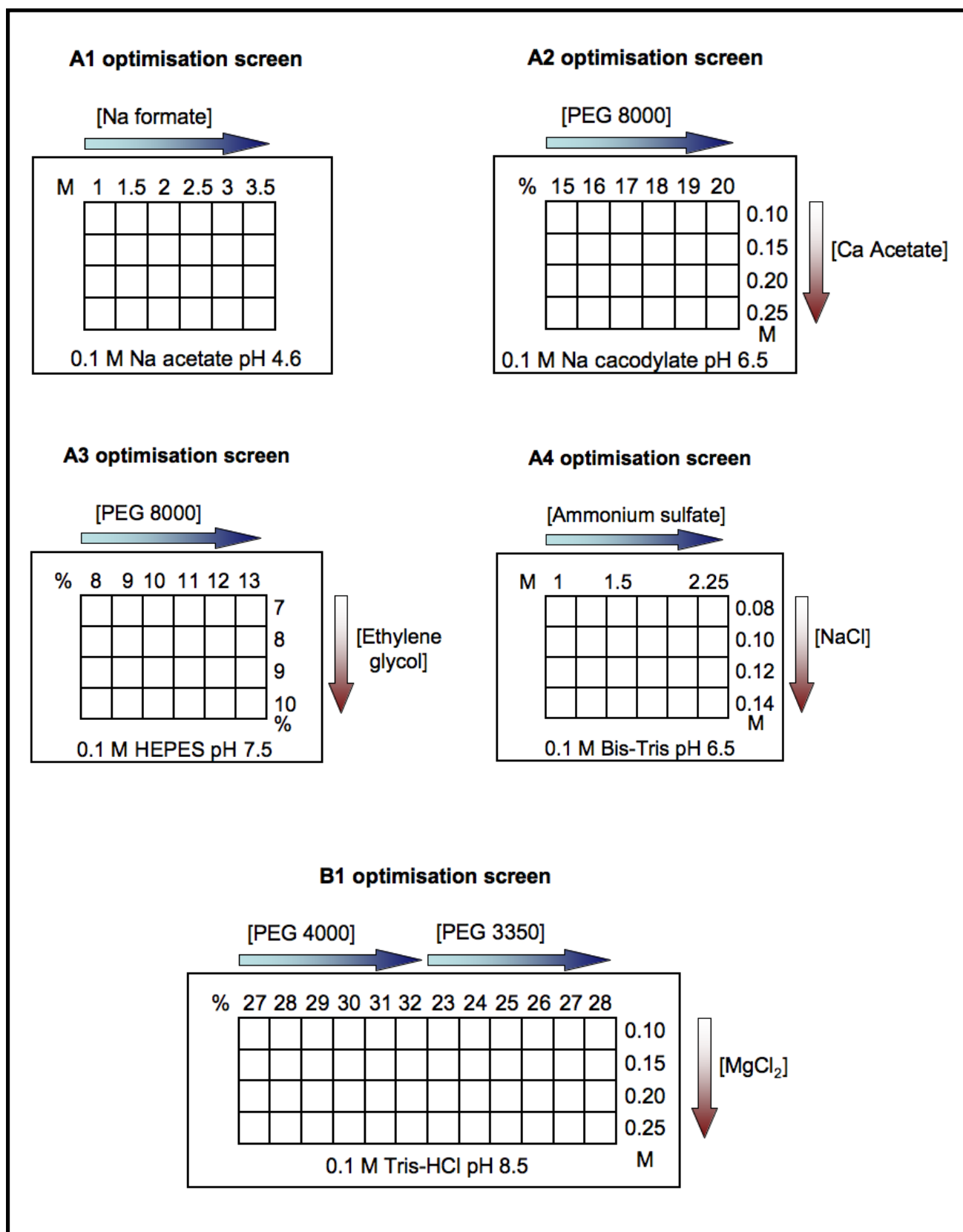


Figure 53. Finer screening around initial crystallisation conditions to find optimum conditions. Self-made optimisation screens were set up for the conditions giving crystals of apo-Tr (conditions A1 to A4) and biotin-Tr (B1).

5.3.2.2 Final crystallisation conditions used for data collection

The final apo-Tr and biotin-Tr crystals used to gather a full data set from the beamline experiment were from the A3 and B1 optimisation screens respectively. The final mother liquor used to grow the apo-Tr crystals was 12 % PEG 8000, 9 % ethylene glycol and 0.1 M HEPES pH 7.5, whilst the final mother liquor used to grow the biotin-Tr crystals was 27 % PEG 4000, 0.25 M MgCl_2 and 0.1 M Tris-HCl pH 8.5. Images of the crystals grown under these conditions are shown in Figure 54.

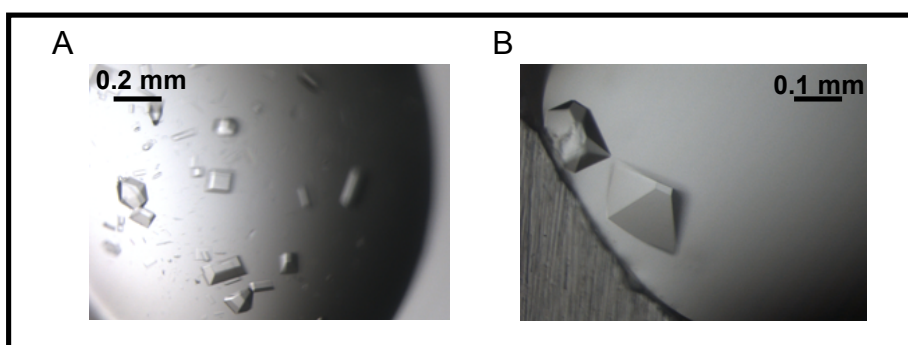


Figure 54. Apo-Tr and biotin-Tr crystals grown after optimisation of crystallisation conditions. Imaged using a Leica MZ16 stereomicroscope. A, Crystals of apo-Tr readily formed. B, Fewer biotin-Tr crystals grew than apo-Tr, but those that did grow were well formed and dispersed.

5.3.2.3 Solution of apo-Tr and biotin-Tr structures

The final data collection and refinement statistics are shown in Table 14. The data collected enabled both the apo-Tr and biotin-Tr structures to be solved at high resolution; apo-Tr was solved at 1.45 Å and biotin-Tr was solved at 1.49 Å.

Molecular replacement (MR) was used to solve both structures, and once solved, the structures were deposited in the PDB with the codes 2Y3E for apo-Tr and 2Y3F for biotin-Tr. An apo-SA structure (PDB code: 1SWB) (Freitag et al., 1997) was used as the search model for the apo-Tr structure, and then the solved apo-Tr structure was used as the search model for the biotin-Tr structure. Before the apo-Tr and biotin-Tr structures could be compared to one another, as well as to other SA structures, it was essential to elucidate whether any model bias had been introduced into the biotin-Tr structure as a result of using the apo-Tr structure as the search model for MR. The use of the calculated phases from the apo-Tr model has the potential to bias the refinement process towards that model, and so to ensure that this was not the case in this instance, a simulated annealing composite omit map was calculated. To do this, a short simulated annealing refinement run was conducted (Brunger, 1991) with a random selection of 10 % of the atoms omitted. The electron-density omit map produced was then aligned to both the apo-Tr and biotin-Tr structures and correlation coefficients calculated (Table 15). The final biotin-Tr structure was found to fit the omit map much better than the apo-Tr structure, indicating the solution of the biotin-Tr structure was not biased by the search model (Figure 55).

	Apo-Tr (PDB: 2Y3E)	Biotin-Tr (PDB: 2Y3F)
Crystallisation	12 % PEG 8000 9 % ethylene glycol 0.1 M HEPES pH 7.5	27 % PEG 4000 0.25 M MgCl ₂ 0.1 M Tris-HCl pH 8.5
Data collection		
Space group	<i>I</i> 4 ₁	<i>P</i> 4 ₂ 2 ₁ 2
Cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	57.59, 57.59, 183.35	57.34, 57.34, 77.55
α , β , γ (°)	90.00, 90.00, 90.00	90.00, 90.00, 90.00
Resolution range (Å)	27.47–1.45 (1.53–1.45)	38.78–1.49 (1.57–1.49)
<i>R</i> _{merge}	0.052 (0.619)	0.069 (0.085)
<i>I</i> / σ <i>I</i>	13.6 (2.0)	26.0 (19.6)
Completeness (%)	97.8 (87.1)	100.0 (100.0)
Multiplicity	3.6 (3.2)	7.0 (7.1)
Refinement		
Resolution range (Å)	27.47–1.45	32.12–1.49
No. reflections	48,884	21,581
<i>R</i> _{work} / <i>R</i> _{free}	0.134 / 0.178	0.135 / 0.152
No. atoms		
Protein	1,855	954
Biotin	n/a	16
Glycerol	18	18
Water	321	142
<i>B</i> factors (Å ²)		
Protein	33.41	10.10
Biotin	n/a	5.83
Glycerol	44.14	28.31
Water	38.19	27.92
rmsd		
Bond length (Å)	0.005	0.009
Bond angle (°)	0.93	1.48
Twinning		
Twin operator	– <i>h</i> , <i>k</i> , – <i>l</i>	n/a
Twin fraction	0.497	n/a

Table 14. Data collection and refinement statistics for apo-Tr and biotin-Tr crystallisation. Values in parenthesis refer to the highest resolution shell. *B* factors are the average of all atoms.

Omit map aligned with:	Correlation coefficient for side chain atoms
Apo-Tr	0.47
Biotin-Tr	0.73

Table 15. Alignment of a simulated annealing composite omit map with the apo-Tr and biotin-Tr structures. The biotin-Tr structure fitted the omit map much better than the apo-Tr structure, revealing the biotin-Tr structure was not biased by the use of the apo-Tr structure as a search model for MR.

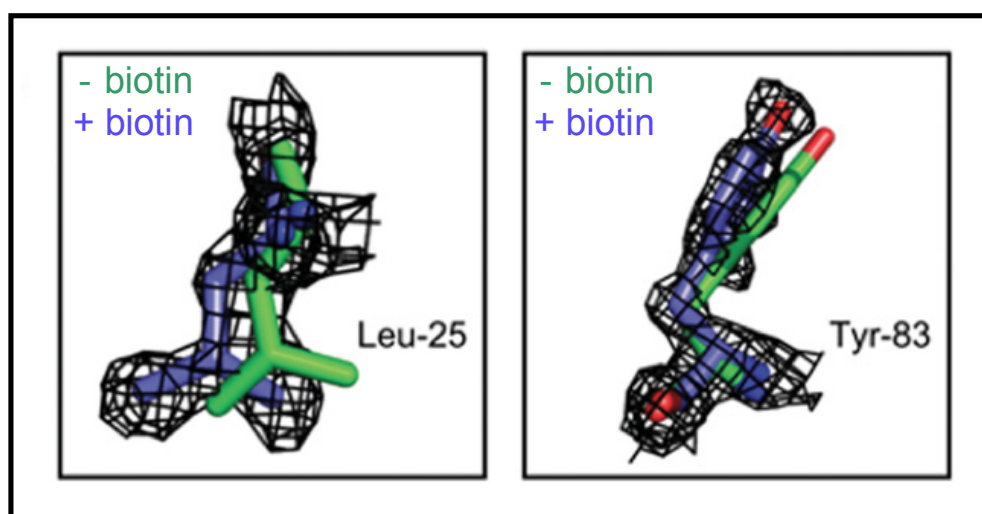


Figure 55. The solution of the biotin-Tr structure was not biased by the search model used for MR. A simulated annealing composite omit map was calculated, using the starting phases from the apo-Tr structure (shown in black mesh, contoured at $1.3\ \sigma$). The omit map was aligned with both the apo-Tr structure (shown in green) and the biotin-Tr structure (shown in blue), with the biotin-Tr structure fitting the omit map much better than the apo-Tr structure. Leu-25 and Tyr-83 are presented here as they clearly illustrate the better fit of the omit map to the biotin-Tr model.

5.3.3 Structural analysis

5.3.3.1 Tertiary and quaternary structure of traptavidin

As there were two Tr subunits in the apo-Tr asymmetric unit and one Tr subunit in the biotin-Tr asymmetric unit, symmetry transformations were applied, to view the complete tetramer (Figure 56).

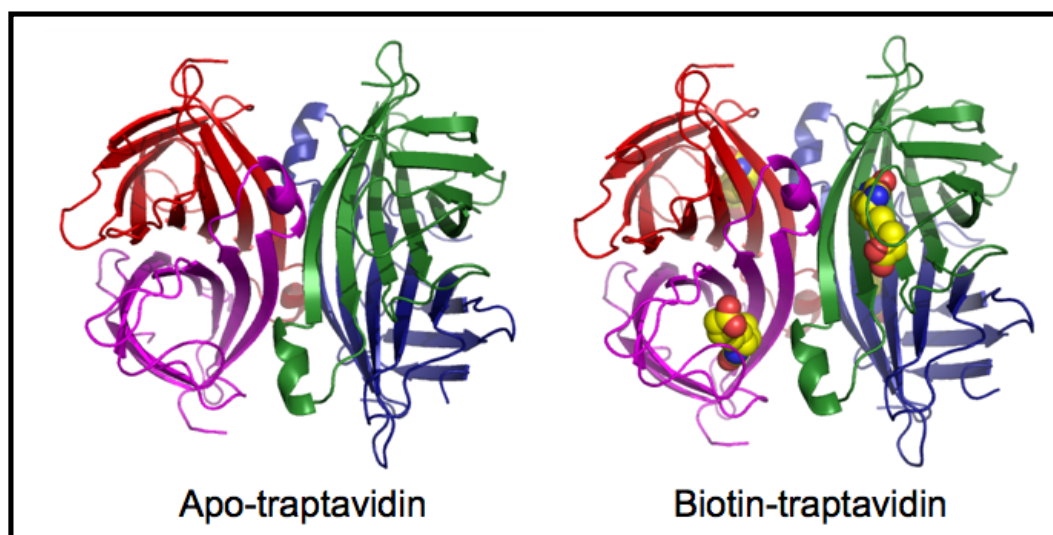


Figure 56. Tetrameric quaternary structure of apo-Tr (left) and biotin-Tr (right). Each Tr subunit is shown in a different colour and the biotin bound to the biotin-Tr structure is shown in space-filling mode.

There was clear and unambiguous electron density for the bound biotin in the biotin-Tr structure (Figure 57).

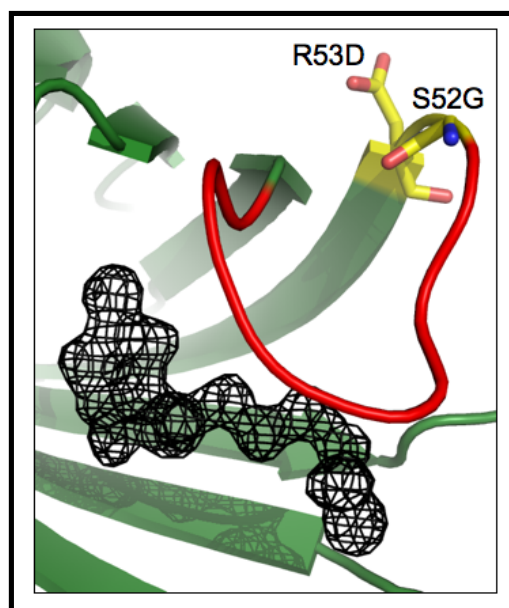


Figure 57. Clear electron density was observed for biotin in the biotin-Tr structure. $F_{\text{obs}} - F_{\text{calc}}$ map of the biotin-binding pocket (shown in black mesh) was overlaid with the protein (shown in cartoon format) and clearly shows the presence of biotin in the pocket. L3/4, which is stabilised in a 'closed' conformation over the bound biotin, is shown in red and the residues that have been mutated to produce Tr are shown in stick format.

It was clear that the overall fold of the individual Tr subunits was very similar to that of SA subunits, with each subunit consisting of eight anti-parallel β -sheets that fold into a β -barrel. Each β -barrel contains a biotin-binding site, located at one end of the barrel, and four subunits come together to give a tetrameric quaternary structure, capable of binding up to four biotin molecules. Backbone alignments of individual subunits of Tr and SA were calculated in order to quantify the similarity of the tertiary structures (Table 16). The two apo-Tr subunits in the asymmetric unit (chains A and B) aligned with the subunits of apo-SA very well, with an average root-mean-square deviation (rmsd) of 0.28 Å and 0.27 Å respectively.

	rmsd (Å) of alignment with apo-Tr (2Y3E)	
	Chain A	Chain B
1SWA chain A	0.30	0.29
chain B	0.29	0.25
chain C	0.27	0.25
chain D	0.27	0.27
Average	0.28	0.27

Table 16. Backbone alignments of the two apo-Tr subunits in the asymmetric unit with apo-SA subunits revealed a high degree of similarity in tertiary structures of apo-Tr and apo-SA. Alignments were calculated using the apo-SA structure 1SWA (Freitag et al., 1997).

The one subunit in the asymmetric unit of the biotin-Tr structure was also aligned with the subunits of the biotin-SA structure 1SWE (Freitag et al., 1997), revealing the high degree of similarity in the tertiary structures (Table 17).

	rmsd (Å) of alignment with biotin-Tr (2Y3F)
1SWE chain A	0.26
chain B	0.26
chain C	0.28
chain D	0.27
Average	0.27

Table 17. Backbone alignments of the biotin-Tr subunit in the asymmetric unit with biotin-SA subunits reveals a high level of similarity in the tertiary structures. Alignments were calculated using the biotin-SA structure 1SWE (Freitag et al., 1997).

5.3.3.2 L3/4 conformation in traptavidin

On visual inspection, the most striking difference seen in the tertiary fold of Tr compared to SA was in the conformation of the loop connecting β -sheets 3 and 4, L3/4. As discussed in Chapter 1, L3/4 plays a critical role in generating the high biotin-binding affinity of SA; it is frequently found disordered in apo-SA in an 'open' conformation, and is stabilised on biotin-binding, forming a 'lid' over the binding pocket in a 'closed' conformation (Freitag et al., 1997). In the apo-Tr structure, however, it was stabilised in a 'closed' conformation over the biotin-binding pocket and remained in this conformation in the biotin-Tr structure (Figure 58).

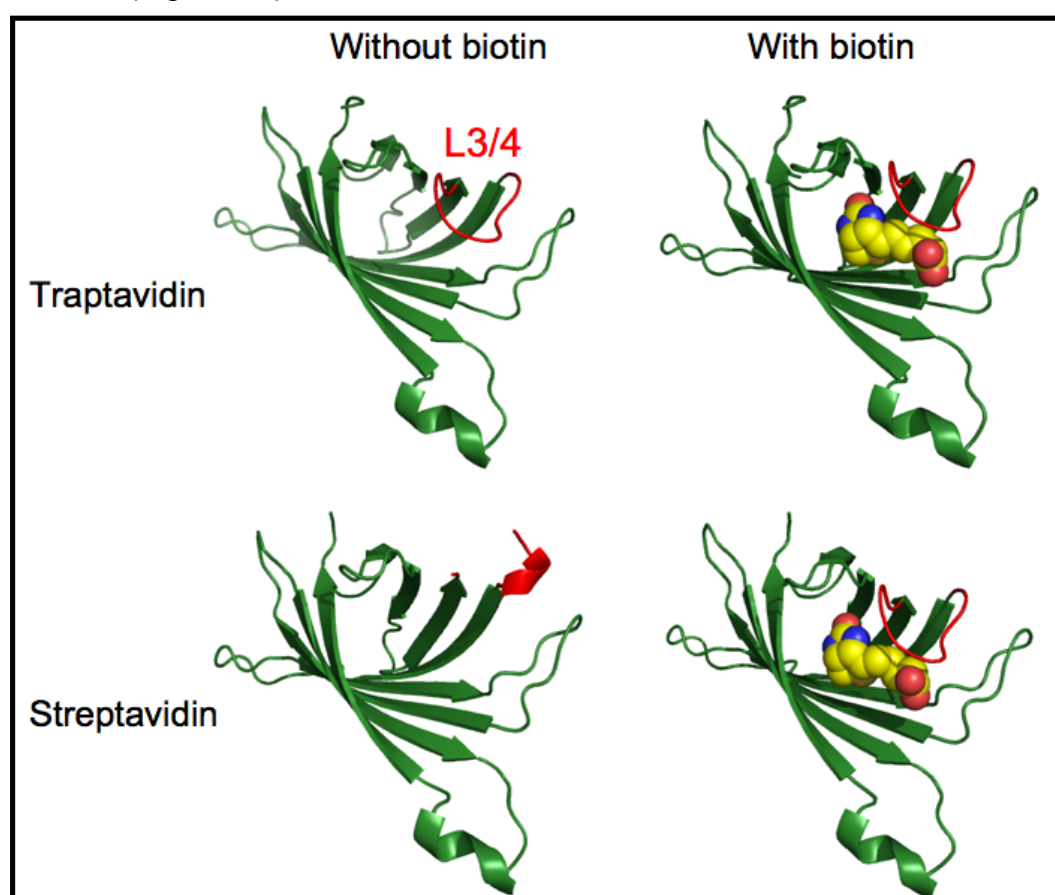


Figure 58. L3/4 is found in a 'closed' conformation in both apo-Tr and biotin-Tr. Comparing tertiary structures of individual subunits of Tr (top) and SA (bottom), L3/4 is undefined in a disordered 'open' conformation in apo-SA (1SWA chain B) but becomes stabilised in a 'closed' conformation over the pocket on biotin-binding (1SWE chain D), whilst it is found in the 'closed' conformation in both apo-Tr and biotin-Tr.

The L3/4 conformation in the apo-SA structures obtained in the physiologically relevant pH range 4.5 – 7.5 was noted. The apo-SA structure 2BC3 had all four L3/4 regions in the tetramer in an ‘open’ conformation (Le Trong et al., 2006), whilst 1SWA and 1SWB had one L3/4 of the four in the tetramer already ordered in a ‘closed’ conformation, and 1SWC had two ordered L3/4 regions (Freitag et al., 1997). Structural alignments of these ordered L3/4 regions with apo-Tr revealed the ‘closed’ loops in 1SWA and 1SWB are comparable to the ‘closed’ L3/4 in apo-Tr (rmsd of 0.24 Å and 0.20 Å respectively), whilst the ordered L3/4 in 1SWC has an entirely different conformation (rmsd of 2.2 Å), resembling an alternative ‘open’ state, with L3/4 projecting away from the biotin-binding site (Table 18).

Apo-TR L3/4 (2Y3E chain A) aligned with ordered L3/4 in:	rmsd (Å)
1SWA chain A	0.24
1SWB chain A	0.20
1SWB chain B	2.2

Table 18. Backbone alignments of the ordered L3/4 regions in different apo-SA structures with apo-Tr L3/4. The ‘closed’ L3/4 in both 1SWA and 1SWB are comparable to the ‘closed’ L3/4 in apo-Tr, whilst the ordered but ‘open’ L3/4 in 1SWC is not comparable.

5.3.3.3 Conformational change on biotin binding

As discussed in Chapter 1, although biotin-binding to SA is not cooperative (Jones and Kurzban, 1995), there is a conformational change in SA on biotin-binding that propagates through the tetramer, involving altered dimer-dimer packing, ordering of surface loops and tighter wrapping of the β -barrels (Weber et al., 1989). To quantify this conformational change, individual subunits of apo-SA and biotin-SA were aligned, and compared to the alignment of apo-Tr and biotin-Tr (Figure 59). The conformational change in Tr on biotin-binding (rmsd of 0.21 Å) is considerably smaller than in SA (rmsd of 0.36 Å).

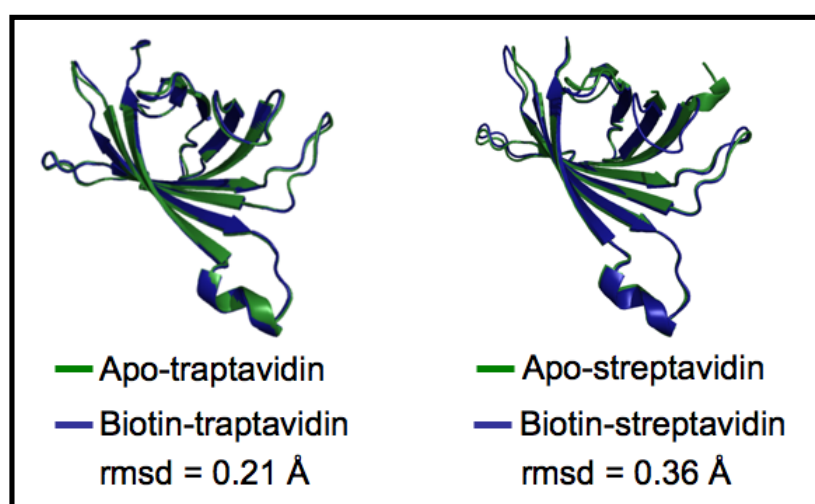


Figure 59. Tr undergoes a smaller conformational change on biotin-binding than SA. Individual subunits of apo- and biotin-bound Tr (left) and SA (right) were aligned and rmsd calculated. The apo-protein is shown in green, with the biotin-bound protein overlaid in blue.

Although the ordering of L3/4 on biotin-binding is well-documented, the conformational change to other loop regions in the protein was investigated, by a series of backbone alignments of the loop regions in the apo- and biotin-bound structures of SA and Tr (Tables 19 and 20). As expected, L3/4 experienced a smaller conformational change on biotin-binding in Tr than in

SA (rmsd of 0.25 Å in Tr compared to 2.6 Å in SA). However, in addition to this, five out of the other six loop regions in Tr experienced a smaller conformational change upon biotin-binding than in SA.

rmsd (Å) of alignment of loop in:			
Apo-Tr aligned with: (2Y3E chain A)	Biotin-Tr (2Y3F chain A)	Apo-SA (1SWA chain B)	Biotin-SA (1SWE chain D)
L1/2 (23-26)	0.050	0.20	0.21
L2/3 (33-38)	0.14	0.23	0.15
L3/4 (45-52)	0.25	2.3 *	0.25
L4/5 (62-70)	0.20	0.20	0.22
L5/6 (79-88)	0.18	0.28	0.18
L6/7 (98-102)	0.14	0.22	0.37
L7/8 (113-121)	0.20	0.23	0.24
Overall subunit	0.21	0.29	0.35

Table 19. Backbone alignment of loop regions in apo-Tr to those in biotin-Tr, apo-SA and biotin-SA. Amino acid residues comprising each loop are shown in parentheses. * The alignment with L3/4 in 1SWA was calculated using only residues 45 and 49-52, as residues 46-48 were disordered and not visible in the structure.

rmsd (Å) of alignment of loop in:			
Apo-SA aligned with: (1SWA chain B)	Biotin-SA (1SWE chain D)	Apo-Tr (2Y3E chain A)	Biotin-Tr (2Y3F chain A)
L1/2 (23-26)	0.27	0.20	0.18
L2/3 (33-38)	0.23	0.23	0.22
L3/4 (45-52)	2.6 *	2.3 *	2.3 *
L4/5 (62-70)	0.25	0.20	0.25
L5/6 (79-88)	0.31	0.28	0.27
L6/7 (98-102)	0.42	0.22	0.26
L7/8 (113-121)	0.17	0.23	0.15
Overall subunit	0.36	0.29	0.25

Table 20. Backbone alignment of loop regions in apo-SA to those in biotin-SA, apo-Tr and biotin-Tr. Amino acid residues comprising each loop are shown in parentheses. * The alignment with L3/4 in 1SWA was calculated using only residues 45 and 49-52, as residues 46-48 were disordered and not visible in the structure.

5.3.3.4 Flexibility analysis of streptavidin and traptavidin

B factor analysis was used to assess the flexibility and dynamics of both apo- and biotin-bound SA and Tr. Analysis was based on relative *B* factors within a structure, rather than making comparisons of absolute *B* factors between different structures, due to the resolution, data collection conditions and refinement method all having an impact on *B* factor (Tronrud, 1996; Yuan et al., 2005b).

The average *B* factor of the main-chain atoms in each residue was plotted for each of the two apo-Tr subunits in the asymmetric unit, to indicate the internal variation in the structure (Figure 60).

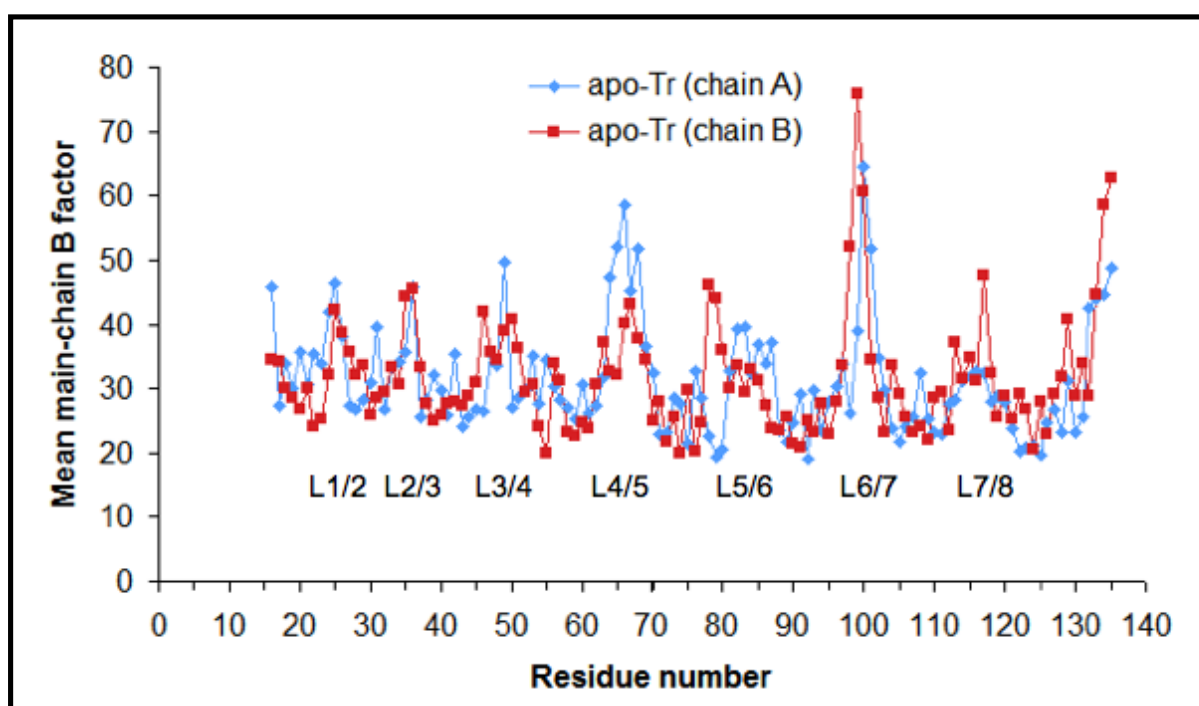


Figure 60. Average main-chain *B* factors for the two apo-Tr chains in the asymmetric unit. Regions corresponding to the loops connecting adjacent β -sheets are indicated.

B factors analysis revealed the β -strand regions between loops had the lowest *B* factor values, indicating they are the least flexible part of the structure, with the peak *B* factor values corresponding to the flexible loop regions, as expected (Figure 61). Overall, the two apo-Tr subunits in the asymmetric unit have very similar patterns of *B* factors distribution, with the main difference being found in L5/6. A possible explanation for this is that one of the chains is involved in crystal contacts at this loop. There is some flexibility seen in L3/4, but obviously not as much as in apo-SA structures, where the residues are frequently found disordered and not visible.

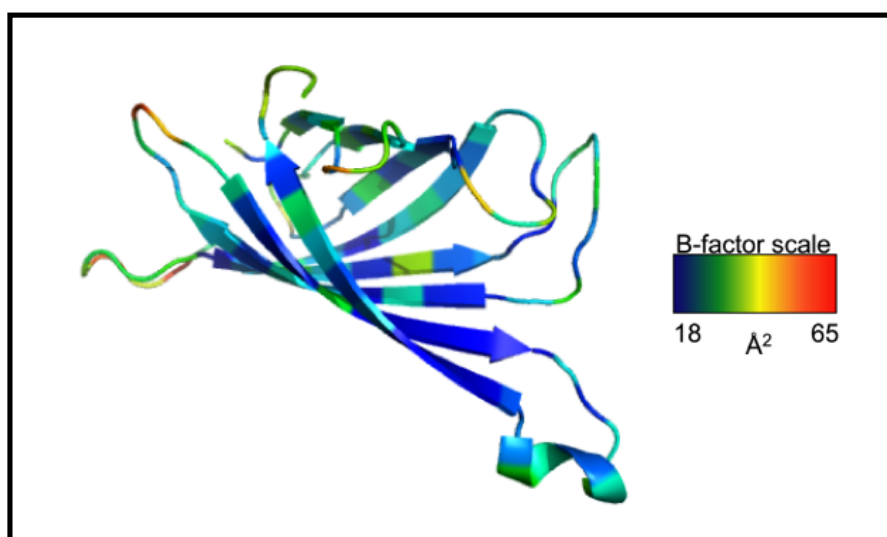


Figure 61. Heat map showing the variation of average *B* factors across the apo-Tr subunit. The red shading corresponds to the most flexible regions, with the highest *B* factors, and the blue shading corresponds to the least flexible regions, with the lowest *B* factors.

When the pattern and distribution of *B* factors across the apo-Tr structure is compared to those in the biotin-Tr structure (Figure 62), the most striking difference is the loss of flexibility of L3/4, which has average *B* factor values almost comparable to the adjacent β -strands in the biotin-Tr structure.

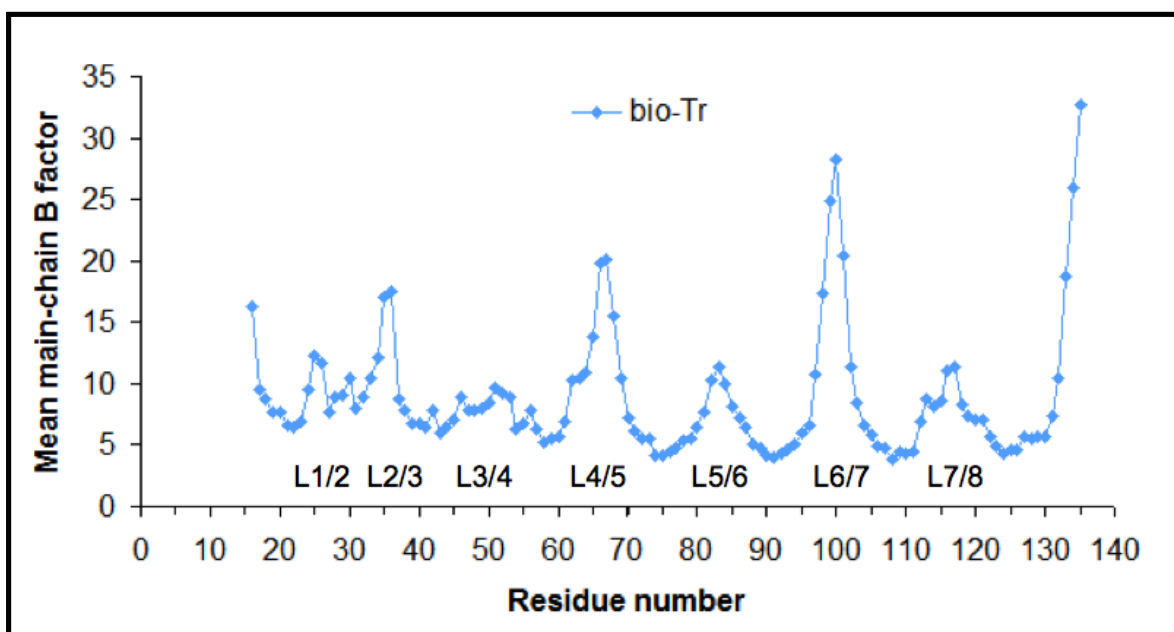


Figure 62. Average main-chain *B* factors for the single biotin-Tr subunit in the asymmetric unit. Regions corresponding to the loops connecting adjacent β -sheets are indicated.

The distribution of B factors across the biotin-Tr structure is similar to that in apo-Tr, with the loop regions having the higher B factors, and therefore higher flexibility, than the β -strands (Figure 63).

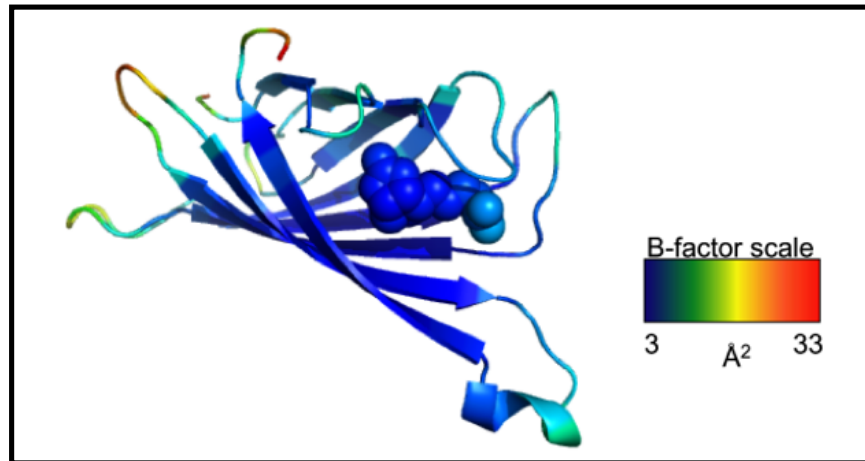


Figure 63. Heat map showing the variation of average B factors across the biotin-Tr subunit. The red shading corresponds to the most flexible regions, with the highest B factors, and the blue shading corresponds to the least flexible regions, with the lowest B factors.

5.3.3.5 Conformation of bound biotin in streptavidin and traptavidin

The conformation of biotin when bound to Tr was investigated, and compared to its conformation when bound to SA. As the ureido oxygen in biotin forms three hydrogen bonds to SA (and Tr), which are arranged with tetrahedral geometry around the ureido oxygen, it has been suggested that this hydrogen-bonding arrangement polarises the ureido group and stabilises a sp^3 oxyanion (DeChancie and Houk, 2007; Weber et al., 1989). The polarisation of this ureido group, resulting in the localisation of negative charge on the oxygen atom and positive charge on the ureido nitrogen atoms, enables biotin to form a greater number of hydrogen bonds, of greater strength, to SA than the ureido group could to solvent molecules (Weber et al., 1992). As a result, the geometry and conformation of biotin is altered when bound to SA / Tr from the structure of biotin in solution (Li et al., 2009). On comparing biotin's bond angles (Figure 64 and Table 21) and bond lengths (Figure 65) when bound to Tr and SA, the geometry was found to be similar in both cases, indicating that changes to biotin's electronic structure when bound to protein cannot account for Tr's specific properties, such as the decreased off-rate.

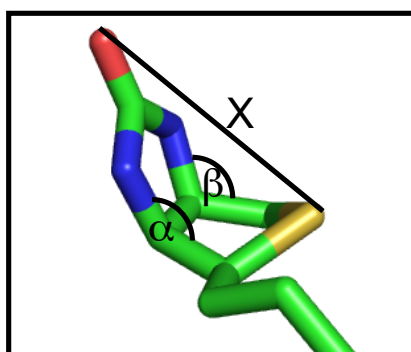


Figure 64. Biotin measurements used to assess geometry when bound to SA and Tr. X represents the distance between the ureido oxygen and sulfur in the tetrahydrothiophene ring, whilst α and β are the angles between the two rings.

Structure: PDB code:	Biotin-SA 1SWD	Biotin-SA 1SWE	Biotin-SA 1MK5	Biotin-Tr 2Y3F
Resolution (Å)	1.90	2.06	1.40	1.49
χ (Å)	4.70	4.70	4.65	4.55
α (°)	114.51	113.94	115.12	114.03
β (°)	114.56	113.92	113.86	110.77

Table 21. Comparing the geometry of biotin when bound to SA and Tr.

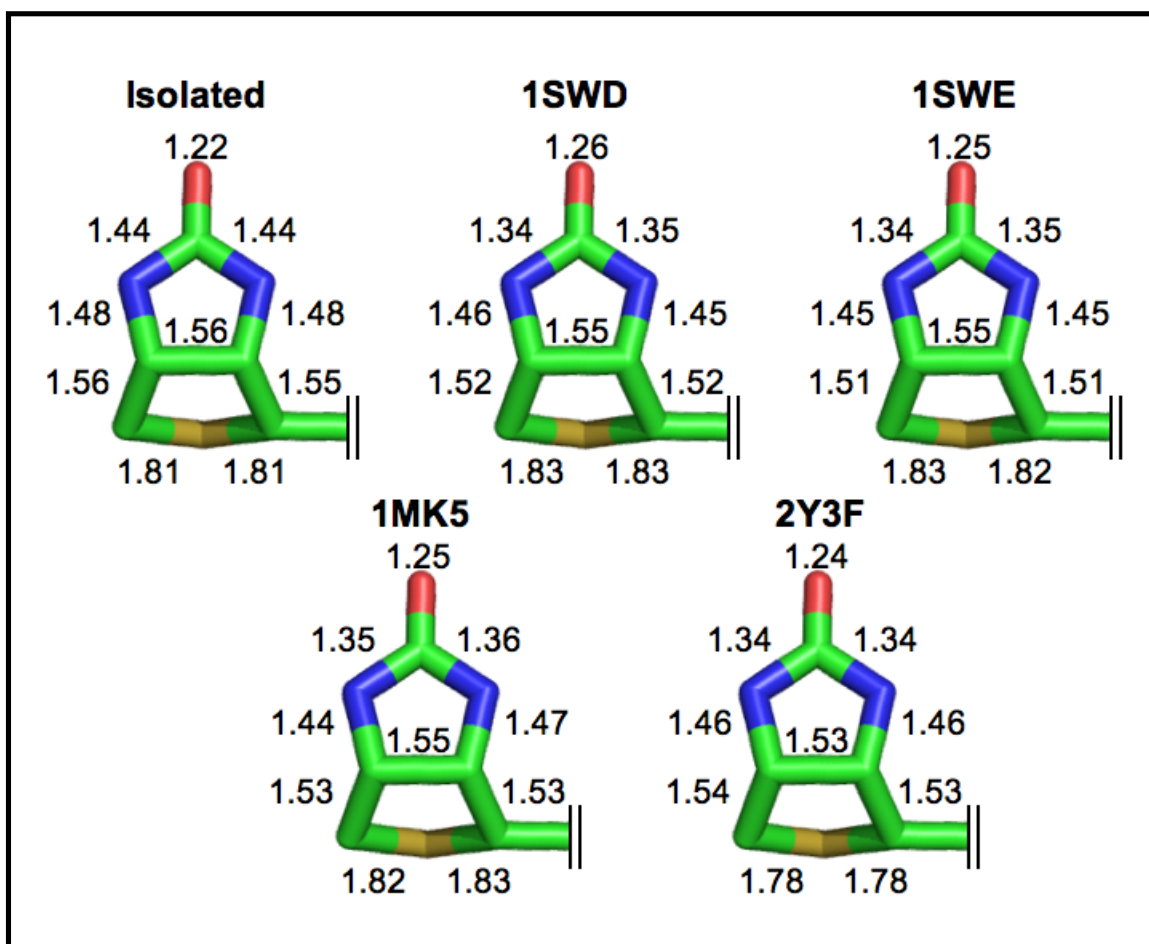


Figure 65. Average bond lengths (in Å) of biotin when in isolation, bound to SA and bound to Tr. Measurements for biotin in isolation taken from Li et al., 2009. Measurements for 1SWD, 1SWE and 1MK5 structures are averaged bond lengths of all biotins in the structures.

5.3.3.6 Hydrogen-bonding network in traptavidin

Due to the importance of the hydrogen-bonding network in generating the high biotin-binding affinity of SA (Weber et al., 1992), the hydrogen bonds formed between Tr and biotin were compared to those in SA-biotin structures, paying close attention to residues Ser-45 and Asn-49, which are located on L3/4, in close proximity to Tr's mutations S52G R53D. The lengths of the hydrogen bonds that Tr makes to biotin are comparable to the SA-biotin hydrogen bonds, when you also consider the variation seen in the different chains of the SA-biotin structures (Figure 66).

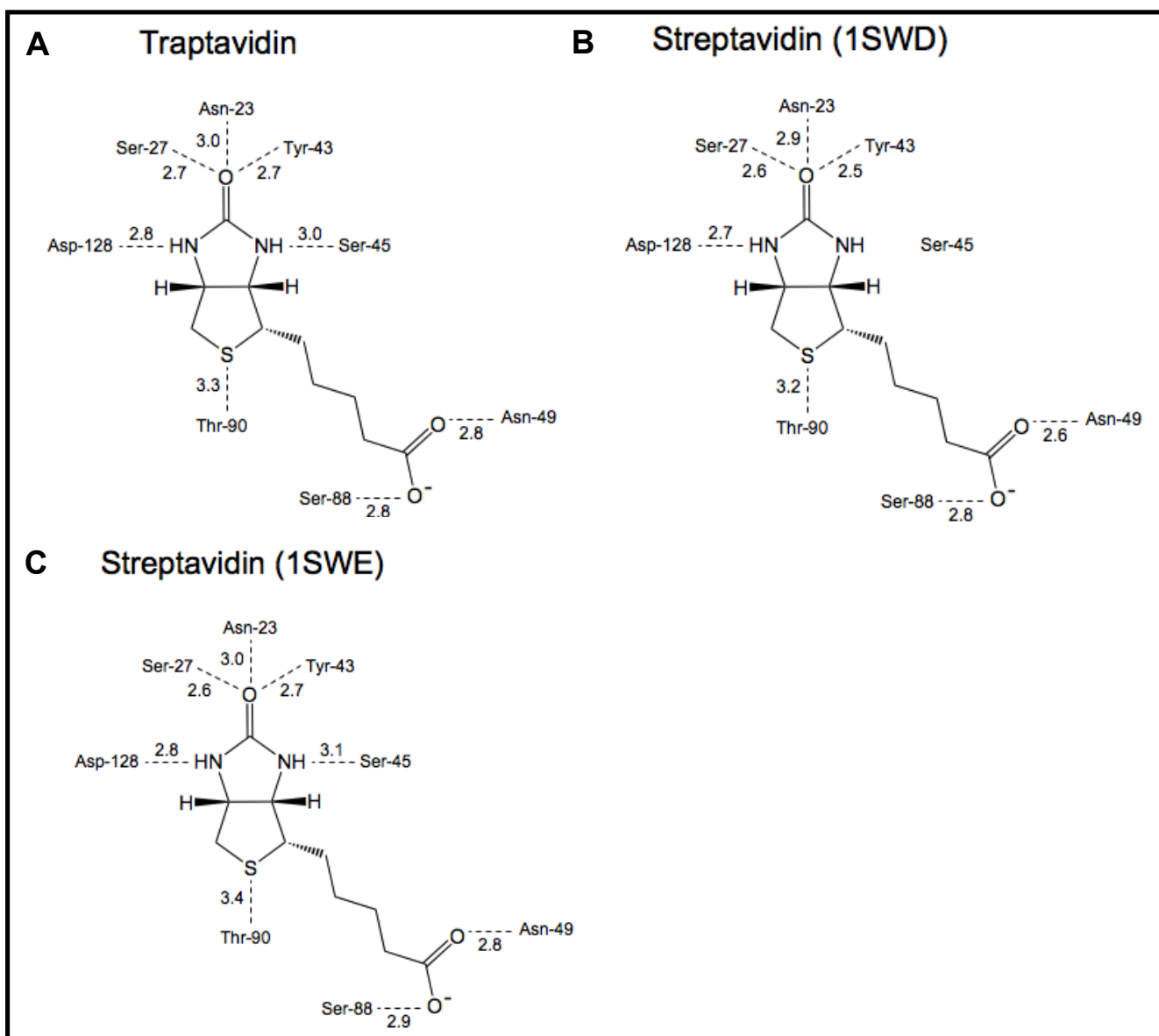


Figure 66. Average hydrogen-bond lengths (in Å) made to biotin in Tr and SA. A. The bond lengths shown for biotin-Tr are the same in all four subunits (as there was one subunit in the asymmetric unit). The dispersion precision indicator (DPI) gave an estimated coordinate error for the biotin-Tr structure of 0.051 Å. B. The hydrogen bond lengths shown for 1SWD are average values for the two biotins in the tetramer. Ser-45 in this structure does not form a hydrogen bond to biotin. The DPI for 1SWD is 0.277 Å. C. The hydrogen bond lengths for 1SWE are average values for the four biotin molecules in the tetramer. The DPI for 1SWE is 0.239 Å.

The presence of the Ser-45 hydrogen bond in only some and not all of the SA-biotin crystal structures alluded to its dynamic nature. For example, in the SA-biotin structure 1SWE, each Ser-45 residue has its side-chain hydroxyl group orientated towards biotin's N3' ureido nitrogen forming a hydrogen bond, whilst in the SA-biotin structure 1SWD, the Ser-45 hydroxyl group is pointed away from the N3' nitrogen and does not form a hydrogen bond (Figure 67). In the biotin-Tr structure, the Ser-45 side-chain hydroxyl group is clearly orientated towards the N3' nitrogen, forming a hydrogen bond 3 Å in length.

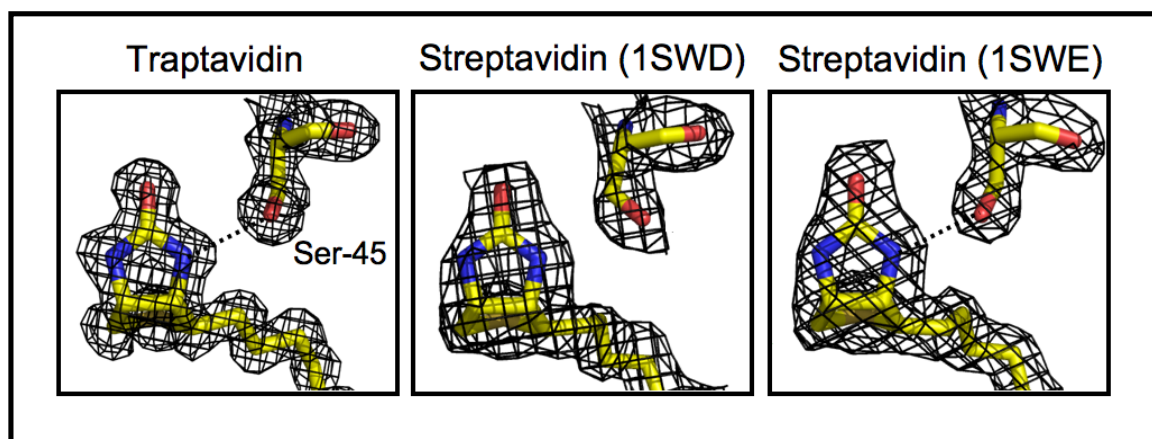


Figure 67. The Ser-45 hydrogen bond to biotin is present in some but not all SA-biotin structures. The electron density for biotin and Ser-45 is shown in black mesh surrounding a stick model of the atoms. Hydrogen bonds are indicated with a dashed line. Left panel: the biotin-Tr structure 2Y3F contains a Ser-45 hydrogen bond to biotin in every subunit. Middle panel: the biotin-SA structure 1SWD does not contain a Ser-45 hydrogen bond. Right panel: the biotin-SA structure 1SWE contains a Ser-45 hydrogen bond in each of the four subunits in the tetramer.

I then focussed my attention on the hydrogen bonding network made by Ser-52 and Arg-53 in wildtype SA, and the mutant Gly-52 and Asp-53 in Tr. Considering Tr, Gly-52 undergoes only a very small conformational change on biotin-binding. Instead, the S52G mutation may play a role in generating Tr's specific properties as the wildtype Ser-52 residue in SA may stabilise an alternative open conformation of L3/4, which in turn would destabilise the Ser-45 hydrogen bond to biotin (Figure 68). In this alternative conformation in SA, the main-chain carbonyl group of Ser-52 forms a hydrogen bond to the main-chain of N-H of Ser-45, whilst the hydroxyl side-chain of Ser-52 makes a hydrogen bond to the main-chain N-H group of Ala-46. This bonding configuration is seen in apo-SA structures with a disordered L3/4 (e.g. 1SWC), but could also transiently form in biotin-bound SA structures, contributing to the rare biotin-dissociation events. As Ser-52 is mutated to Gly in Tr, the potential for this alternate, rival hydrogen bonding configuration to form when biotin is bound is removed, thereby making biotin dissociation even more unlikely.

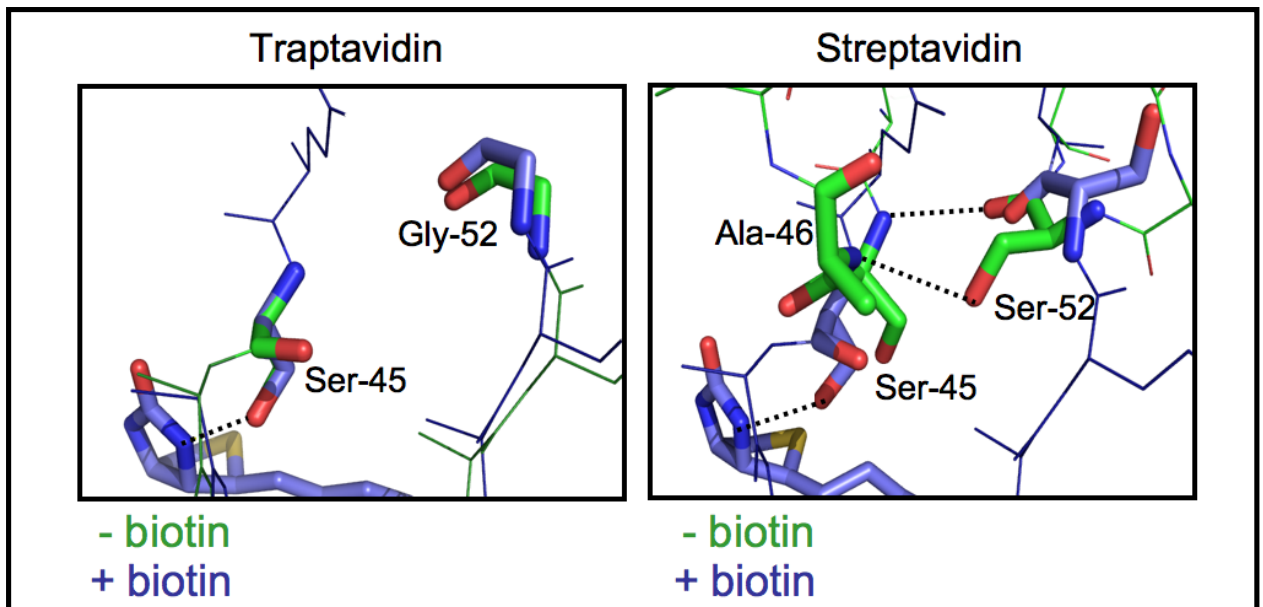


Figure 68. Alternative, rival hydrogen-bonding by Ser-45 in SA may contribute to the rare events that lead to biotin dissociation. Overlays of the L3/4 region in both apo- and biotin-bound proteins, with the relevant residues displayed in stick format and hydrogen bonds shown as dashed lines. Left panel: Both apo-Tr and biotin-Tr have an ordered 'closed' L3/4, with no hydrogen bond between Ser-45 and the mutant Gly-52. Right panel: In apo-SA with a disordered and undefined L3/4 (1SWC chain B), Ser-52 forms hydrogen bonds with both Ser-45 and Ala-46. On biotin-binding (1SWE chain D), L3/4 is stabilised in a 'closed' conformation and Ser-45 now forms a hydrogen bond to biotin.

5.3.3.7 Intersubunit contacts

In order to investigate whether the properties of Tr arose from altered intersubunit contacts within the tetramer, monovalent Tr (Tr1D3) was made, consisting of one Tr subunit and three 'dead' (D) SA subunits (Figure 69). The 'dead' SA subunits contain the mutations N23A, S27D and S45A, which renders the protein unable to bind biotin (Howarth et al., 2006).

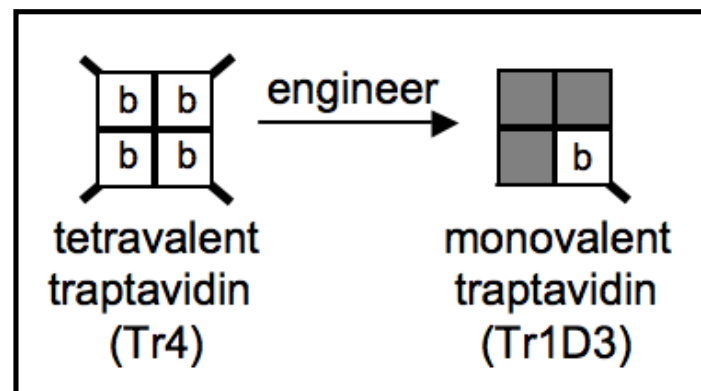


Figure 69. Monovalent Tr (Tr1D3) contains one Tr subunit and three 'dead' SA subunits, incapable of binding biotin. Tetravalent Tr (shown on left), with four Tr subunits each with a His₆-tag (bold diagonal line), is capable of binding four molecules of biotin (b). Monovalent Tr (shown on right) contains only one Tr subunit and three 'dead' SA subunits, and can therefore only bind one molecule of biotin.

A biotin-4-fluorescein (B4F) off-rate assay was conducted on SA, Tr and Tr1D3 and revealed that Tr1D3 had an equivalent off-rate to tetravalent Tr (Figure 70). This indicated that the improved biotin-binding stability of Tr arises from within the individual subunits themselves and not from any altered intersubunit interactions.

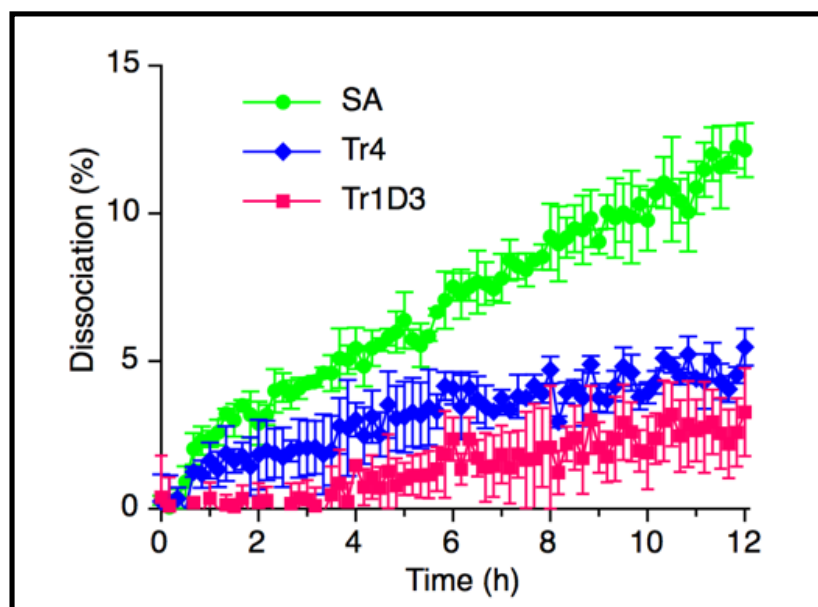


Figure 70. Monovalent Tr had a comparable B4F off-rate to tetravalent Tr. Both T1D3 and Tr4 had a decreased B4F off-rate compared to SA, when measured at 37 °C over 12 hours. Means of triplicate readings are shown $\pm$ 1 s.d. Data collected by A. Koner.

On comparison with SA, Tr was found to have higher tetramer thermal stability, as well as increased biotin-binding thermal stability (Chapter 4), as Tr remained tetrameric and biotin-bound at higher temperatures. Here, the thermal stability of Tr1D3 was investigated using a thermostability assay and SDS-PAGE, to visualise the dissociation of tetramer into both Tr and D subunits (Figure 71).

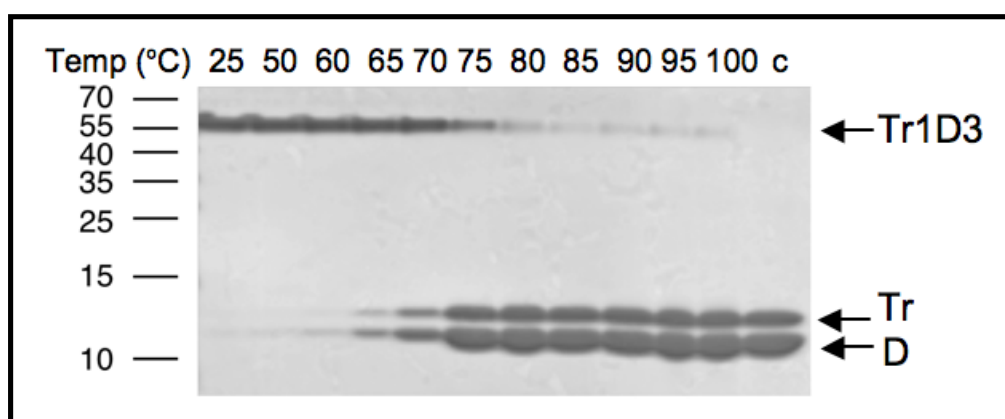


Figure 71. Thermostability of Tr1D3. Tr1D3 was incubated at the temperatures indicated for 3 minutes and then analysed by SDS-PAGE, to reveal dissociation into Tr and D monomers (positions indicated). The 100% monomer positive control (c) was heated in 6X SDS loading buffer at 95 °C for 3 min prior to loading.

Tr1D3 thermostability was then compared to that of Tr and SA (Figure 72). Monovalent SA (with one wildtype SA subunit and three 'dead' subunits) has been shown to have comparable thermal stability to tetrameric SA (and also tetrameric 'dead' SA, D4) (Howarth et al., 2006). The temperature at which half the Tr4 protein dissociated to monomer was approximately 10 °C higher than for SA / D4 (see also Chapter 4). However, the thermostability of Tr1D3 was comparable to that of D4 (and SA), as a result of the less stable 'dead' subunits dominating the thermostability of the tetramer.

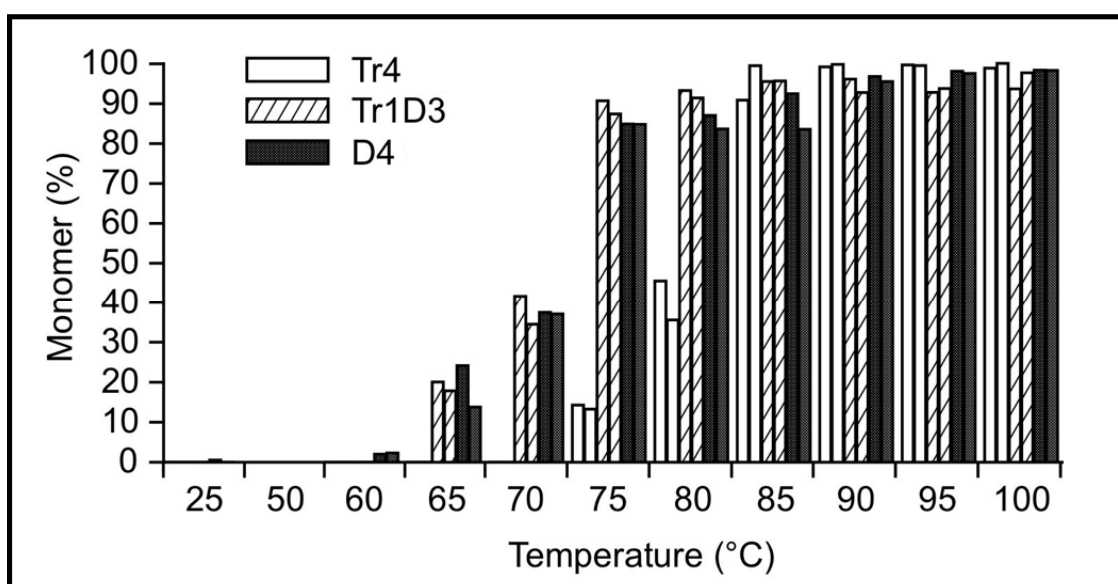


Figure 72. Tr1D3 has comparable thermostability to D4. Tr, D4 and Tr1D3 samples were heated at the indicated temperature for 3 minutes followed by SDS-PAGE, revealing the dissociation into monomers. The percentage monomer from duplicate gels was then plotted.

5.4 Discussion

It was hypothesised in Chapter 4 that the S52G R53D mutations in Tr may have an effect on the flexibility of L3/4, due to the region's critical role in generating the high biotin-binding affinity of SA and the proximity of the mutations in Tr to L3/4. The crystal structures presented in this chapter have confirmed that this is the case, as well as illuminating other key differences between SA and Tr: the numerous structural alignments between apo- and biotin-bound SA and Tr have shown that Tr undergoes a smaller conformational change on biotin-binding, suggesting Tr's structure is pre-formed for biotin-binding, in addition to suggesting that the Ser-45 hydrogen bond to biotin in Tr is more stable than in SA. However, the conformation of bound biotin and the hydrogen bond lengths to biotin in SA and Tr were found to be comparable.

The stabilisation of L3/4 into a 'closed' conformation on biotin-binding is a common feature within the avidin superfamily. In avidin, the L3/4 region is three residues longer than in SA, but still becomes stabilised into a 'closed' conformation on biotin-binding; however, L3/4 remains disordered when biotin-conjugates are bound, explaining avidin's lower affinity for biotin-conjugates when compared to SA (Pazy et al., 2002). Tamavidin (Takakura et al., 2009), AVR4 protein (Eisenberg-Domovich et al., 2005) and rhizavidin (Meir et al., 2009) all have an ordered L3/4 in a 'closed' conformation when biotin is bound. However, it is interesting to note that AVR4 has a 'closed' L3/4 in its apo-form (PDB 1Y53), like Tr. This increased rigidity of L3/4,

probably brought about by a proline residue in L3/4 that is not present in Av or SA, results in an increase in the thermostability of apo-AVR4 (T_m of 106 °C) compared to apo-Av (T_m of 84 °C) (Eisenberg-Domovich et al., 2005). Also interesting to note is that rhizavidin, a naturally occurring dimeric protein, has its L3/4 highly restrained by a disulfide bond linking it to L5/6, resulting in L3/4 being stabilised in a 'closed' conformation in both apo- and biotin-bound rhizavidin (Meir et al., 2009). Rhizavidin, due to its dimeric quaternary structure, lacks a conserved Trp-120 in the biotin-binding pocket that has been shown to be crucial for generating SA's high-affinity binding (Sano and Cantor, 1995). However, the increased rigidity of L3/4 is thought to compensate for the lack of this tryptophan residue as rhizavidin is still capable of high-affinity biotin and biotin-conjugate binding (the K_d of 2-iminobiotin binding is 2.2×10^{-7} M for rhizavidin and 1×10^{-7} M for Av) (Meir et al., 2009). In a similar way, bradavidin II, a tetrameric biotin-binding protein from *Bradyrhizobium japonicum*, lacks the conserved tryptophan crucial for high-affinity biotin-binding, but yet is capable of high-affinity binding as its L3/4 region is highly constrained due to the presence of a proline residue and a disulfide bond in L3/4 linking this loop to L5/6 (Helppolainen et al., 2008).

As well as acting as a 'lid' against biotin dissociation and contributing to the high binding affinity, L3/4 also plays a role in biotinyl *p*-nitrophenyl ester (BNP) hydrolysis (see Section 1.2.3.1), which some members of the avidin superfamily are capable of carrying out. The disordered L3/4 in Av is thought to aid hydrolysis by increasing the solvent accessibility of BNP when bound, whilst the constrained and ordered L3/4 in AVR4 is thought to impose

conformational restrictions on BNP, making BNP more vulnerable to nucleophilic attack (Prizant et al., 2006). Indeed, when L3/4 of AVR4 was transferred into Av, there was a marked increase in the ability of the resultant chimeric avidin to hydrolyse BNP (Hayouka et al., 2008).

The decreased flexibility of L3/4 in apo-Tr explains the decreased on-rate of biotin and biotin-conjugates. However, *B* factor analysis revealed L3/4 still possesses some flexibility (the average *B* factors for this loop region are considerably higher than the adjacent β -strands), and it is this flexibility that results in an acceptable biotin on-rate. It is reasonable to assume that if L3/4 did not have this flexibility in the apo-Tr structure, the on-rate would be too low for most practical applications of Tr. *B* factor analysis also revealed the remarkably low *B* factors for L3/4 in the biotin-Tr structure (L3/4 values were comparable to those of the adjacent inflexible β -strands), which explains the extremely low off-rate of Tr. Both hydrogen/deuterium-exchange (Williams et al., 2003) and isothermal titration calorimetry (Weber et al., 1992) have shown the decreased flexibility of SA on biotin-binding, and it would be very interesting to use these techniques to further investigate the change in Tr flexibility on biotin-binding. In particular, hydrogen/deuterium exchange analysis at a single residue resolution (Rand et al., 2009) would be extremely informative.

The altered hydrogen-bonding network in Tr suggested a possible mechanism for the S52G R53D mutations in Tr to exert their effect on L3/4 flexibility. In apo-SA with a disordered L3/4, Ser-52 is found hydrogen-bonded across the

base of the loop region to Ser-45 (as well as to Ala-46). On biotin-binding, this Ser-52 : Ser-45 hydrogen-bonded interaction is lost, which results in L3/4 becoming stabilised in a 'closed' conformation and Ser-45 forming a hydrogen bond to biotin instead. In Tr, the S52G mutation removes the possibility of hydrogen bonding to Ser-45, and instead, L3/4 is found in a 'closed' conformation in both apo-Tr and biotin-Tr.

Previous crystallographic, biophysical and computational research into the biotin-SA dissociation pathway suggested that biotin dissociation was initiated by a water molecule that enters the binding pocket and competes with the Asp-128 hydrogen bond for binding to an ureido nitrogen in biotin (Freitag et al., 1999; Grubmuller et al., 1996; Hyre et al., 2002). However, in all wildtype SA-biotin crystal structures available, the Asp-128 hydrogen bond to biotin is always visible, whilst it is the Ser-45 hydrogen bond that is frequently absent. Therefore, it is likely that the breakage of the Ser-45 hydrogen bond is the first event on the biotin dissociation pathway. Once the Ser-45 hydrogen bond to biotin is lost, an alternative hydrogen-bonding scheme can then arise in SA, as Ser-45 is capable of bonding to Ser-52 at the other side of the base of L3/4. This hydrogen-bonding scheme can then cause a conformational change in L3/4, favouring loop 'opening' and biotin dissociation. The S52G mutation in Tr removes the possibility of this alternative hydrogen-bonding scheme forming across the base of L3/4, making the rare fluctuations that lead to L3/4 'opening', and hence biotin dissociation, even less likely to occur. This is supported by the very recent publication of the crystal structure of a SA mutant containing the single mutation S52G, in which L3/4 is found stabilised

in a 'closed' conformation in both the apo- and ligand-bound forms (Magalhaes et al., 2011). As Arg-53 does not directly interact with biotin, one could speculate that the R53D mutation in Tr aids the S52G mutation by further reducing L3/4 flexibility.

It has been found that thermophilic proteins have a tendency to have shorter exposed loop regions than their mesophilic homologues, which acts to lower the entropy of unfolding (Thompson and Eisenberg, 1999). Also, many proteins have been shown to have an increased thermal stability when their conformational flexibility has been reduced and the probability of partial unfolding events occurring is minimised (Corazza et al., 2006; Rader, 2009; Wray et al., 1999). Although SA derives from mesophilic bacteria, it is a highly thermostable protein, and Tr has even higher thermostability. The ~ 10 °C increase in thermostability of Tr compared to SA (Chapter 4) could contribute to the increased biotin-binding stability of Tr by reducing the frequency of thermally-induced conformational fluxes that could disrupt protein structure and lead to dissociation. However, the analysis of the biotin-binding stability and thermostability of monovalent Tr revealed that the increased binding stability of Tr is not a result of altered intersubunit interactions. The demonstration that monovalent Tr has comparable biotin-conjugate binding stability to tetravalent Tr means that monovalent Tr has the potential to replace SA (and Tr) in applications where tetravalency causes problems, such as in single-molecule imaging where tetravalent SA (or Tr) would give cross-linking (Howarth et al., 2006; Howarth et al., 2008). Also, divalent and trivalent Tr can be made in the same manner as SA2D2 and

SA3D1 (Howarth et al., 2006) and these species are anticipated to have comparable binding stability to Tr (and Tr1D3), thereby further increasing the utility of Tr to applications where it is necessary to have proteins of defined valency, such as in nanoassembly (Kesapragada et al., 2010).

Chapter 6: Conclusions

6.1 A summary of the results of this work

The research presented here has shown that it has been possible to rationally design a SA mutant capable of more stable biotin-binding than wildtype. This was previously regarded as a challenging goal, due to the assumption that SA's biotin-binding had been optimised by nature and the existence of a large number of SA mutants (over 200) that had impaired biotin-binding ability. However, this work highlights that, for SA at least, it is physically possible to increase the binding stability past that which is biologically relevant (and hence that which has been selected for by evolution). Taking a rational mutagenesis approach in this work was time-consuming and rather risky; indeed, the 19 mutants detailed in Appendix A1 were designed to have more stable biotin-binding than wildtype SA, but on production and characterisation were found to have lower biotin and biotin-conjugate binding affinity. These weaker-binding SA mutants exhibited a wide range of affinities, from only slightly impaired biotin-binding ability compared to SA, through to dramatically reduced binding ability. However, these weaker mutants are anticipated to still be of use in the laboratory, for example, in generating a calibration curve to dissect force generation by motor proteins (Croizat et al., 2010), and enabling easier purification of biotinylated molecules of interest by affinity chromatography. It should be noted, however, that using a directed evolution method, which is commonly regarded as a robust method in protein engineering (Romero and Arnold, 2009), would have been difficult as the

high-affinity biotin-binding would have made selection difficult. Also, selecting for binding to a particular biotin-conjugate may have resulted in an increased binding affinity towards that molecule only, and not a general increased binding stability. Furthermore, there was a large variation in the yields of the eight higher-stability mutants after refolding, with some mutants being more susceptible to misfolding than others. These poorly-folding mutants may have had a decreased likelihood of being present in a library in a directed evolution approach (Bloom et al., 2006).

Traptavidin, one of the eight SA mutants produced in this study that had a decreased dissociation rate from B4F (presented in Chapter 3), was characterised in Chapters 4 and 5. The characterisation in Chapter 4 revealed that in addition to a decreased biotin and biotin-conjugate off-rate, Tr's biotin-binding had increased mechanical and thermal stability (Chivers et al., 2010). However, it was also found that Tr had a decreased biotin and biotin-conjugate on-rate. Although in *Streptomyces avidinii* it is assumed that evolution would favour a SA with a rapid on-rate and slow off-rate, in many laboratory applications the off-rate is of higher importance than the on-rate. In these cases, such as when SA is used as a molecular anchor on biosensor surfaces or arrays, the use of Tr would be advantageous over SA. The increased thermal and mechanical stability of Tr would also be advantageous in these applications (in addition to the applications outlined in Chapter 4) as the increased thermostability of Tr would mean that arrays / sensor chips would have a longer shelf-life, particularly when stored at room temperature, which is particularly relevant to point-of-care biosensors being used in

developing countries, where there is frequently no cold-storage facilities for the sensor chips / arrays. Also, the increased mechanical stability would result in less dissociation from the array / sensor as a result of fluid flow across the surface (Pierres et al., 2002).

The crystal structures presented in Chapter 5 helped shed light on the origins of the increased biotin-binding stability in Tr and the role played by the S52G R53D mutations in Tr. The S52G mutation, found in the second shell of hydrogen-bonding residues in the binding pocket, alters the hydrogen-bonding network, which affects L3/4 conformation (Chivers et al., 2011; Magalhaes et al., 2011). The R53D mutation, however, is neither in the binding-pocket nor in the hydrogen-bonding network, and so it is hypothesised that R53D acts to further reduce the flexibility of L3/4. This rather nebulous role of the R53D mutation is not unexpected, however, as it is already known that mutations distant from a protein's binding site can still have an effect on function (Shimotohno et al., 2001).

It is hoped that the work detailed in this study will be built upon in future to produce a family of SA mutants, having varied characteristics (such as altered kinetics, specificity, valency, thermal and mechanical stability), but with each member being ideally suited to its particular application, and together forming a SA 'toolbox' for biotechnology.

6.2 Streptavidin's future prospects

Although SA and Tr both possess characteristics that make them very good biotin-binding proteins (such as a rapid on-rate, slow off-rate and low non-specific binding), there are opportunities for improving them even further. Previous work by others has involved reducing SA's non-specific *in vivo* binding (Murray et al., 2002; Wilbur et al., 2004) and immunogenicity (Marshall et al., 1996), in addition to achieving controlled biotin binding and release (Ding et al., 2001; Stayton et al., 1999). I undertook some provisional work towards increasing the utility of SA / Tr, and Appendix A2 outlines efforts towards increasing the ease and yield of recombinant expression, which is a major consideration if a protein is to become a widely-used biotechnological tool.

As well as producing novel SA mutants with a variety of characteristics to expand the utility of the SA-biotin interaction, it is hoped that future research will address two key questions: firstly, whether it is possible to produce a SA monomer possessing the characteristic biotin-binding stability of wildtype SA, or even Tr; and secondly, whether an increased biotin-binding stability and decreased off-rate will always be accompanied by a decrease in on-rate. Mutation of the interface residues in the SA tetramer has produced monomeric SA (detailed in Chapter 1) but the loss of the conserved Trp-120 (which is donated by the neighbouring subunit in the tetramer) means the biotin-binding pocket is incomplete in monomeric SA, dramatically lowering the biotin-binding affinity (Wu and Wong, 2005, 2006). Provisional experiments have suggested it could be possible to 'rescue' the biotin-binding

affinity of monomeric SA by incorporation of the Tr mutations S52G R53D, and it is hoped that future research will investigate the feasibility of this. In support of this, when the S52G mutation was introduced into a mutant SA protein which had increased desthiobiotin-binding stability but decreased biotin-binding stability, the new S52G-containing mutant had higher biotin-binding stability, increasing the biotin half-life six-fold (Levy and Ellington, 2008). Also, a survey of other members of the avidin superfamily (detailed in Chapters 1 and 5) reveals that rhizavidin is capable of high-affinity biotin-binding despite being a dimer and lacking the conserved binding-pocket residue Trp-120 (Helppolainen et al., 2007). This has been attributed to L3/4 being physically constrained by a disulfide bond, which is thought to compensate for the lack of the conserved tryptophan residue (Meir et al., 2009). It would be very interesting, therefore, to investigate whether restraining the conformation of a SA mutant monomer with disulfide bonds would reduce the conformational fluxes that lead to biotin dissociation, thereby increasing the stability of biotin-binding. It is possible that a conformationally-restricted monomer would have a binding pocket that was less accessible to biotin and hence a decreased on-rate, but it would be interesting to investigate whether compensating mutations could be introduced in the periphery of the binding pocket to, for example, increase the on-rate through electrostatic steering. Increasing the electrostatic attraction between an antibody fragment and its antigen by mutation of antibody residues has already been used to increase the on-rate of the interaction (Marvin and Lowman, 2003).

I have also conducted some provisional work into expressing SA and Tr as single-chain dimers, and it is hoped this work will be continued in the future. Having two, and ultimately four subunits, joined in one polypeptide chain will increase the utility of SA as a scaffold, as the subunits can then be independently mutated to generate a completely defined tetramer, which would, for example, enable genetic encoding of controlled valency tetramers. Currently, the introduction of a mutation into the primary sequence of SA results in the mutation being present in all four subunits, or, if a mixed-refold method is used to generate the protein (Howarth et al., 2006), a mixture of homo- and hetero- mutant tetramers. The use of intersubunit linkers to create the single-chain dimer / tetramer could also increase the stability of the resultant protein, by reducing conformational flux in the tetramer. This has already been demonstrated in a higher-stability SA mutant with intersubunit crosslinks (Reznik et al., 1996). A single-chain tetravalent avidin has already been produced (Nordlund et al., 2005a), which should encourage SA engineering efforts to produce a spatially and stoichiometrically defined tetramer.

Although there is already sufficient knowledge to enable computational (Das and Baker, 2008) and structure-guided (Holm et al., 2009) design of protein-ligand interactions at the micromolar and nanomolar level, the factors responsible for generating picomolar and femtomolar affinities are less well documented (Midelfort et al., 2004). Indeed, the mutations in Tr responsible for its increased biotin-binding stability are found not in the first shell of binding-site residues directly hydrogen-bonded to biotin, but in the second

shell of residues (hydrogen-bonded to the residues in the first shell), which contributes to the difficulty in optimising this high-affinity interaction computationally. It is hoped that the production and characterisation of Tr presented in this thesis, in addition to future characterisation by isothermal titration calorimetry and hydrogen/deuterium-exchange with a single-residue resolution, will contribute to illuminating the molecular origins of extreme-affinity interactions. This would have implications for the rational design of high-affinity protein interactions, such as the design of drugs with improved efficacy.

Appendix A1: Biotin-4-fluorescein binding data for all streptavidin mutants made in this study

As a rapid and facile initial screen of all SA mutants produced, the biotin-4-fluorescein (B4F) off-rate was measured at 37 °C over 10 hours (detailed method given in Section 3.2.3) to visualise biotin-conjugate binding stability. A total of 27 SA mutant proteins were made in this study using rational site-directed mutagenesis. These can be subdivided into eight mutants with higher stability biotin-conjugate binding than wildtype SA (which were characterised further in Chapters 3, 4 and 5 of this thesis) (Figure 73), five mutants with a comparable or slightly impaired binding ability (Figure 74), and 14 mutants with a greatly decreased biotin-conjugate binding ability (Figure 75). A summary of the B4F data and brief explanation of the rationale behind the production of the mutants is given in Table 22.

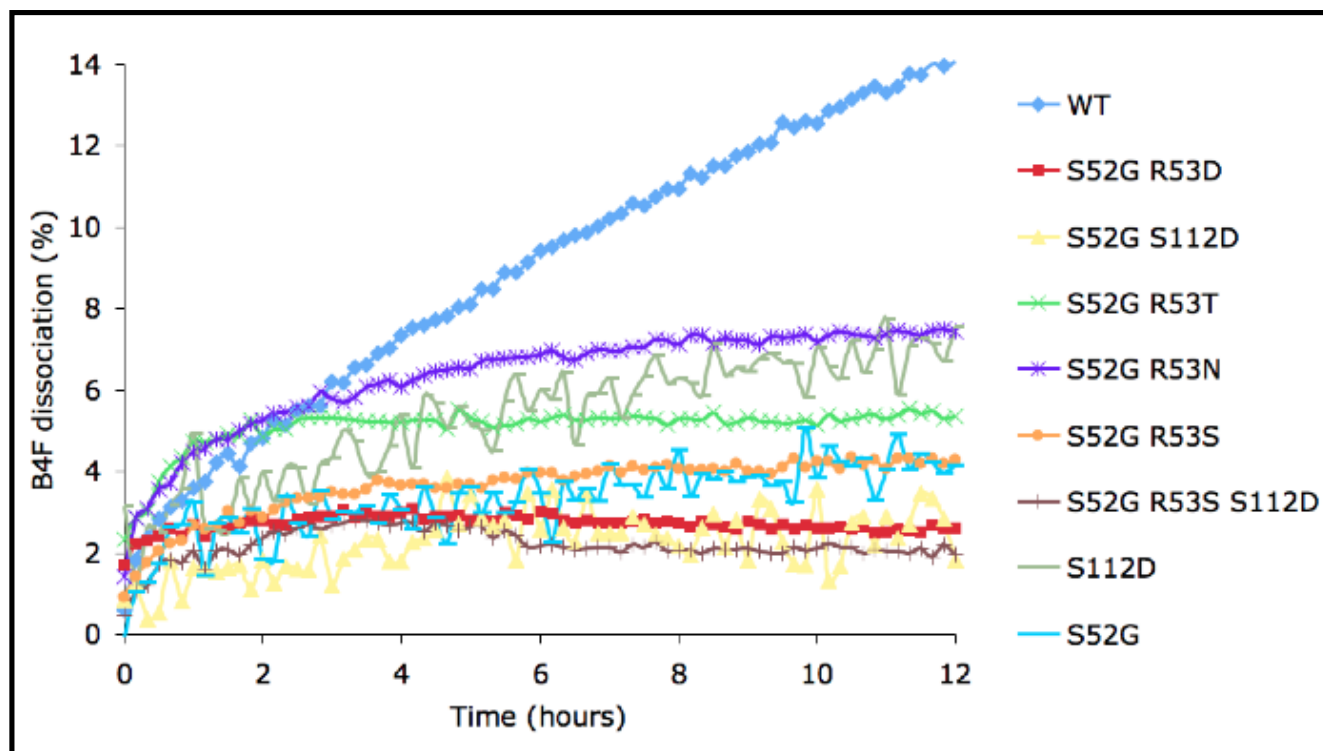


Figure 73. B4F off-rate data for the eight higher-stability SA mutants and wildtype SA. Means of triplicate experiments plotted. For clarity, error bars have been omitted (see Chapter 3 for data with error bars). Assay conducted in PBS with 0.12 mg/mL BSA at 37 °C.

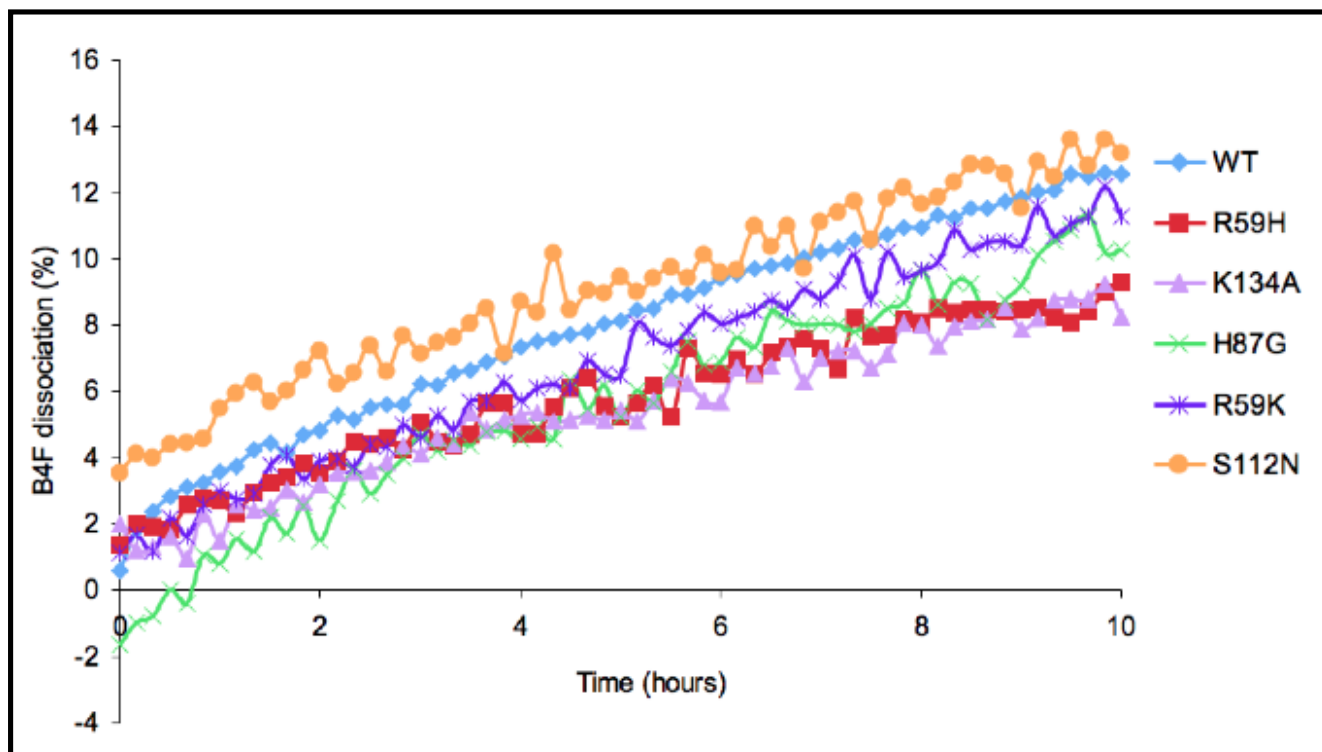


Figure 74. B4F off-rate data for the five SA mutants with comparable binding-stability to wildtype SA. Means of triplicate experiments plotted. For clarity, error bars have been omitted, but all were less than 2 %. Assay conducted in PBS with 0.12 mg/mL BSA at 37 °C.

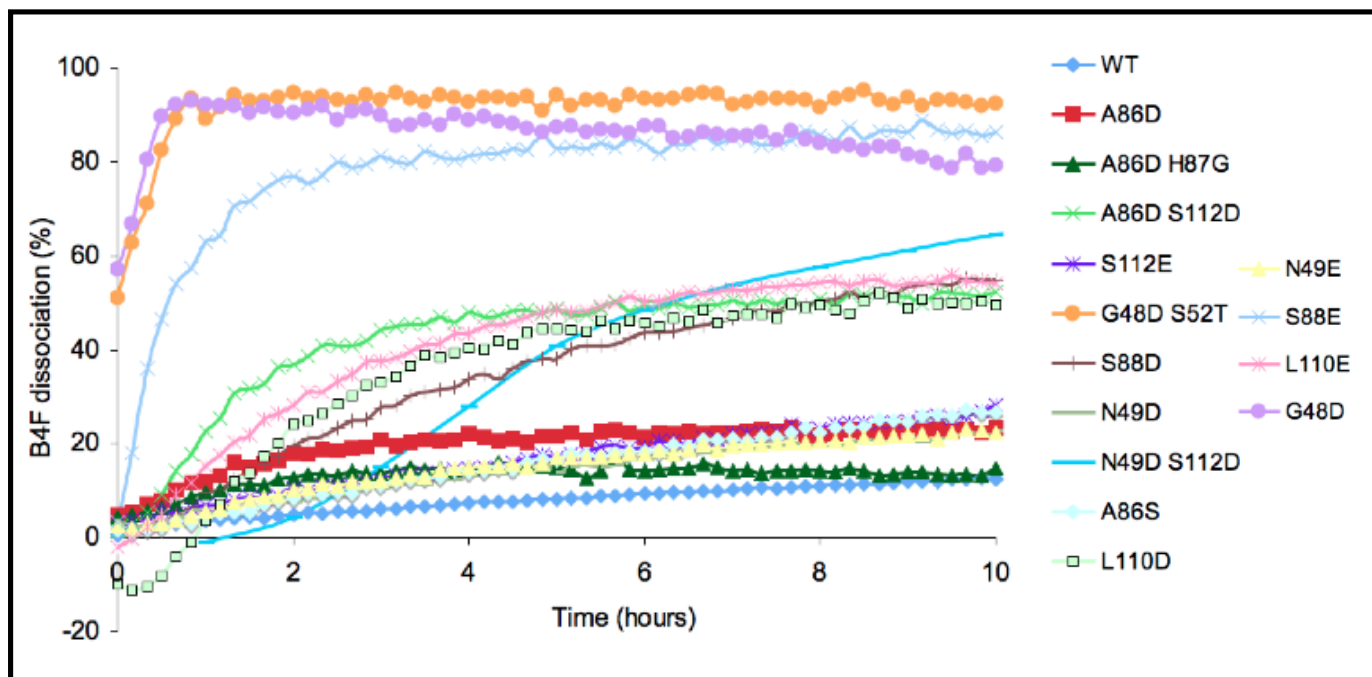


Figure 75. B4F off-rate data for the fourteen weaker-stability SA mutants and wildtype SA. Means of triplicate experiments plotted. For clarity, error bars have been omitted, but were all less than 6 %. Assay conducted in PBS with 0.12 mg/mL BSA at 37 °C.

Protein	Rationale for mutations	Dissociation of B4F after 10 hr at 37 °C (% $\pm$ 1 s.d.)
Wildtype SA	n/a	12.6 $\pm$ 0.6
Higher-stability mutants		
S52G	Proximity to L3/4 & biotin	3.9 $\pm$ 0.8
S112D	Proximity to biotin's valeryl group	7.1 $\pm$ 2.2
S52G R53D	Proximity to L3/4 & biotin	2.6 $\pm$ 2.1
S52G R53N	Proximity to L3/4 & biotin	7.2 $\pm$ 1.1
S52G R53S	Proximity to L3/4 & biotin	4.3 $\pm$ 0.2
S52G R53T	Proximity to L3/4 & biotin	5.1 $\pm$ 0.1
S52G S112D	Proximity to L3/4 & biotin	3.6 $\pm$ 0.9
S52G R53S S112D	Proximity to L3/4 & biotin	2.2 $\pm$ 0.7
Comparable stability to wt SA		
R59H	Possible role in cell binding	9.3 $\pm$ 0.5
R59K	Possible role in cell binding	11.3 $\pm$ 1.8
H87G	On L5/6; close to biotin valeryl	10.3 $\pm$ 0.5
S112N	Proximity to biotin's valeryl group	13.2 $\pm$ 1.6
K134A	Possible role in cell binding	8.3 $\pm$ 0.4
Lower-stability mutants		
G48D	On L3/4; proximity to biotin	79.3 $\pm$ 1.5
G48D S52T	On L3/4; proximity to biotin	92.3 $\pm$ 5.9
N49D	Asn-49 H-bonded to biotin; on L3/4	22.9 $\pm$ 0.3
N49D S112D	On L3/4; close to biotin valeryl	85.2 $\pm$ 0.8
N49E	Asn-49 H-bonded to biotin; on L3/4	22.5 $\pm$ 0.9
A86D	On L5/6; close to biotin valeryl	23.3 $\pm$ 0.9
A86D H87G	On L5/6; close to biotin valeryl	14.7 $\pm$ 0.9
A86D S112D	On L5/6; close to biotin valeryl	52.4 $\pm$ 1.3
A86S	On L5/6; close to biotin valeryl	26.7 $\pm$ 0.5
S88D	Ser-88 H-bonded to biotin	54.8 $\pm$ 2.7
S88E	Ser-88 H-bonded to biotin	86.3 $\pm$ 1.4
L110D	Close to biotin's valeryl group	49.5 $\pm$ 3.1
L110E	Close to biotin's valeryl group	54.0 $\pm$ 1.3
S112E	Proximity to L7/8 & biotin	28.1 $\pm$ 1.0

Table 22. Summary of B4F dissociation data and rationale for making each mutant. B4F dissociation was measured at 37 °C over 10 hours using the method described in Chapter 3. Mean percentage dissociation of triplicate experiments shown $\pm$ 1 s.d. L3/4, loop connecting β -strands 3 and 4. L5/6, loop connecting β -strands 5 and 6. L7/8, loop connecting β -strands 7 and 8.

Appendix A2: Optimising streptavidin / traptavidin expression and purification protocols

A2.1 Introduction

A2.1.1 Streptavidin expression systems

SA is most frequently recombinantly expressed in bacteria, with *E. coli* being the favoured species (Miksch et al., 2008; Sano and Cantor, 1990; Sano et al., 1995; Sorensen et al., 2003; Thompson and Weber, 1993). This can be explained by the ease with which one can genetically manipulate *E. coli*, as well as the low cost and ease of large-scale culturing of the bacteria and its rapid protein expression (Francis and Page, 2010). The most commonly encountered bacterial SA expression systems are detailed below. However, it should be noted that SA has been expressed in other host cells, including: the Gram-positive bacteria *Bacillus subtilis* (Nagarajan et al., 1993); the native producer *Streptomyces avidinii* (Kolomiets and Zdor, 1998), which is often used in large-scale industrial expression systems; baculovirus-infected Sf9 insect cells (Laitinen et al., 1999); and tobacco and apple plants (Markwick et al., 2003).

A2.1.1.1 Expression in bacteria

'Core' SA (residues 13-139), consisting of SA with truncated N and C termini, is the favoured construct used for expression in bacteria (Pahler et al., 1987). The full-length protein has reduced biotin-binding ability (Bayer et al., 1989; Pahler et al., 1987), possibly due to residues at the C terminus folding back and occupying the biotin-binding site (Le Trong et al., 2006), as well as a higher tendency to aggregate (Sano et al., 1995). The most common expression system for SA used in a research setting, recombinant expression of core-SA in *E. coli*, leads to the accumulation of SA as insoluble aggregates (inclusion bodies) in the cytoplasm of the cells (Sano and Cantor, 1990). This is advantageous, as when sequestered into inclusion bodies, SA cannot bind biotin, an essential vitamin needed for cell survival. If SA remained soluble within the cell and able to bind biotin and biotinylated enzymes, SA expression would be toxic to the host cell. However, insoluble expression of SA into inclusion bodies does have its disadvantages. In order to purify SA, inclusion bodies have to be solubilised with a strong denaturant, commonly with 6 M guanidine hydrochloride at low pH (pH ~ 1.5), which also denatures the SA contained within them (Howarth and Ting, 2008; Sano and Cantor, 1990). SA then has to be refolded by removing the denaturant, commonly either by rapidly reducing the denaturant concentration by dilution into refolding buffer, or slowly reducing the denaturant concentration by dialysis (Vallejo and Rinas, 2004). This process is very time- and labour-intensive, with the success of the refolding step having a considerable impact on the final protein yield. For these reasons, it is desirable to have an expression system that gives soluble expression of SA in bacterial cells.

A2.1.1.1.1 Soluble expression in bacteria

As *E. coli* expression of SA in most expression systems results in the formation of inclusion bodies, the Gram-positive *B. subtilis* was investigated as an alternative host cell for recombinant expression (Nagarajan et al., 1993). Using this expression system, yields of 20 mg/L were achieved. However, expensive biotin-free minimal media was needed for the cultures, which were typically grown for long time periods (~ 12 hours). The SA protein produced in this way contained a N-terminal signal peptide for secretion into the growth medium and was a mixture of different length SA species, due to partial processing at the N and C termini. However, this *B. subtilis* expression system utilising a SA N-terminal signal peptide has recently been used to achieve soluble expression of SA mutants (Wu and Wong, 2005).

The native producer *S. avidinii* has been utilised for soluble SA expression, particularly in commercial settings, as it can give very high SA concentrations (Kolomiets and Zdor, 1998). However, the productivity of this system is very low compared to *E. coli*, as a result of the more complex genetic manipulation protocols for *S. avidinii*, as well as its longer generation time (Payne et al., 1990).

In *E. coli* BL21 (DE3) pLysS cells, soluble expression of SA has been achieved using a cold shock vector expression system, giving a yield of 10 mg/L (Matsumoto et al., 2011). However, it should be noted that the SA being expressed had a pentapeptide tag at the C terminal, and the yield achieved is

comparable to the yields reported earlier in this work of insoluble higher-affinity SA mutants purified by refolding.

The 12 residue T7 tag has also been fused to the N-terminal of core-SA in order to increase its solubility when expressed in *E. coli* (Gallizia et al., 1998). This system was reported to give yields of 70 mg/L, but others failed to reproduce this, instead finding leaky expression prior to induction led to cell death (Humbert et al., 2008). The T7 tag has also been utilised, together with highly optimised and tightly controlled expression conditions (using a fermentor), to give impressive yields of 100-120 mg/L (Humbert et al., 2008). However, the complex experimental set-up and tightly-regulated expression conditions may not encourage researchers to attempt to express and purify SA themselves in this way, although this method should still be attractive to commercial companies who are already familiar with large-scale protein expression systems.

A very simple soluble SA expression system giving impressive yields was reported by Sorensen and coworkers (Sorensen et al., 2003). They investigated the effect of different solubility partners on the soluble yield of SA, testing the 40 kDa maltose-binding protein (MBP), the 55 kDa transcription anti-termination factor protein NusA and three permutations of the N-terminal region of initiation factor 2 protein (IF2) as N-terminal fusions for His₆-tagged SA. MBP (Kapust and Waugh, 1999) and NusA (Davis et al., 1999) have previously been shown to promote correct folding and greatly increase the solubility of recombinant proteins expressed in *E. coli*, but Sorensen et al. were the first to demonstrate

the use of IF2 as a solubility partner. The N-terminal region of *E. coli* IF2 protein is composed of three domains: domain I (residues 1 to 158), which has been structurally characterised by NMR (Laursen et al., 2003), is thought to account for the high solubility of the IF2 protein and has some resistance to SDS denaturation (as it does not migrate according to its weight on SDS-PAGE); domain II (residues 159 to 290), which has a less ordered structure and also has some resistance to SDS denaturation; and domain III (residues 291-390), which also has a less ordered structure (Sorensen et al., 2003). The various SA-fusion proteins were overexpressed in both BL21 (DE3) cells (expression at 37 °C for 3 hours) and C41 (DE3) cells (expression at 20 °C for 5 hours) and the expression data is shown in Table 23.

Fusion Partner	Molecular weight (kDa)	Soluble expression in:					
		BL21 at 37 °C for 3 hr:			C41 at 20 °C for 5 hr:		
		%	mg	nmol	%	mg	nmol
MBP	229.6	82	58	253	67	25	109
NusA	284	84	290	1021	81	15	53
IF2 I to III	237.6	29	91	383	74	74	311
IF2 I to II	193.6	40	102	527	68	89	460
IF2 I	114.8	47	107	932	86	96	836

Table 23. Expression data from SDS-PAGE densitometry data (Sorensen et al., 2003). The figures refer to 1 L of bacterial culture or 3-5 g wet weight of cells. Permission obtained (Copyright Clearance Centre License number: 2751301252690).

The soluble fractions from expression trials of the five solubility partners shown in Table 23 were also analysed for ease of isolation using His₆ affinity chromatography and biotin-binding ability (Figure 76).

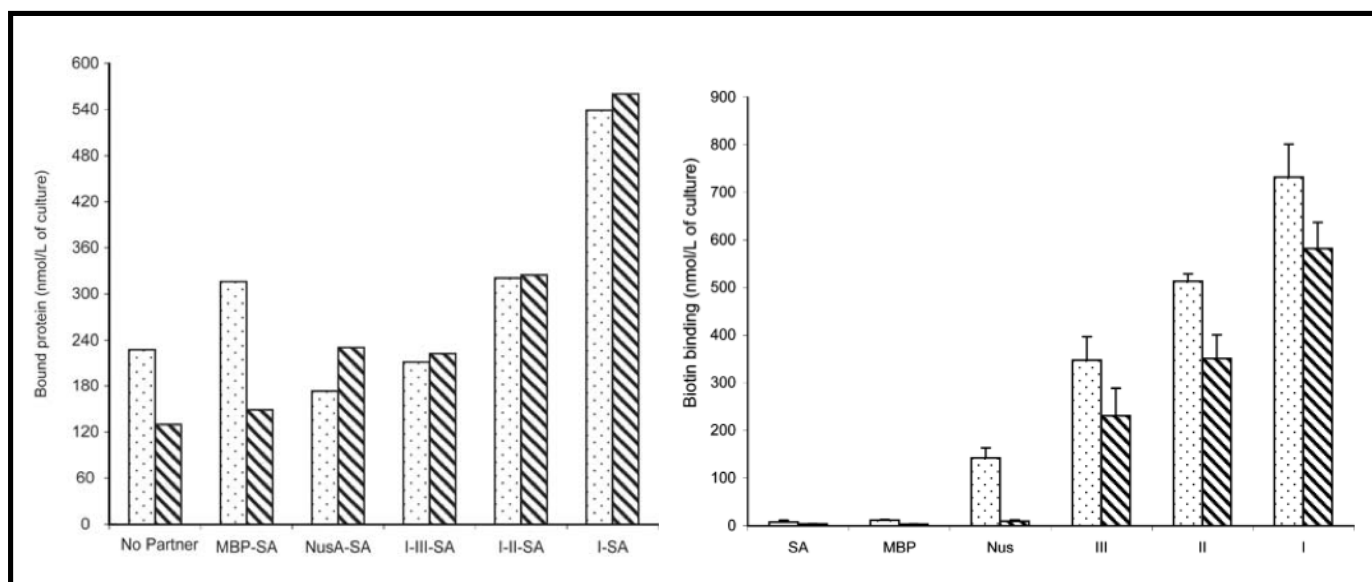


Figure 76. Isolation of protein (left) and biotin-binding ability (right) of the soluble fraction from BL21 (DE3) cells (dots) and C41 (DE3) cells (hashes) expressing various SA-fusion proteins (Sorensen et al., 2003). Permission obtained (Copyright Clearance Centre License number: 2751301252690).

These analyses revealed that there were some discrepancies between the expression levels indicated by SDS-PAGE densitometry data and the amount of active, soluble protein that could be purified. For instance, the NusA fusion partner generated very favourable expression data, but could not be purified by nickel immobilised metal affinity chromatography (IMAC) chromatography and had low biotin-binding ability, indicating that a large amount of the soluble NusA-SA protein was inactive. The researchers concluded that the most favourable SA solubility partner is domain I of IF2 (Sorensen et al., 2003).

A2.1.2 Aim of this chapter

For a protein to become a standard tool in biotechnology, as is hoped will be the case for Tr, scientists should be able to obtain large quantities of the protein cheaply and easily, either by producing it themselves or purchasing it from a commercial source. The current method for expression of Tr involves insoluble expression in *E. coli*, leading to accumulation at high concentration in inclusion bodies within the bacteria. The purification of protein from these inclusion bodies results in high-purity preparations with little contaminating proteins, but the protocol is extremely time-consuming and labour-intensive.

Soluble bacterial expression would result in a more rapid and facile production of large quantities of Tr, and indeed is the method used for commercial, large scale SA production. Therefore, the feasibility of using soluble Tr expression and a simple purification protocol was investigated in this chapter. The plasmid encoding the IF2 domain I fused to SA was acquired from K. Mortensen, and expression and purification trials of IF2-SA were undertaken. Site-directed mutagenesis was then used to introduce the Tr mutations S52G R53D into IF2-SA, producing IF2-Tr. Extensive expression and purification trials of IF2-Tr were then undertaken, finally resulting in the production of core-Tr.

A2.2 Materials and methods

A2.2.1 IF2-SA plasmid

The pET-15b-I-SA plasmid was a kind gift from K. Mortensen, University of Aarhus. This plasmid encoded domain I of Initiation Factor 2 (IF2) (158 aa) fused to the N terminal of SA, and is hereafter called IF2-SA. The sequence was confirmed by sequencing with T7F and T7R primers (see Table 24).

A2.2.2 Expression trials of IF2-SA

The IF2-SA plasmid was subcloned into three *E. coli* expression strains according to the manufacturer's protocol (see Table 25 for details): BL21 (DE3) RIPL cells (Agilent Technologies, UK); BL21 (DE3) pLysS cells (Agilent Technologies, UK); and B834 (DE3) pLysS cells (EMD Chemicals, UK). Transformants were selected on 2x TY agar plates (1.6 % tryptone, 1 % yeast extract, 0.5 % NaCl, 1.5 % agar) supplemented with 100 µg/mL ampicillin (Amp) and 25 µg/mL chloramphenicol (Cam). Individual freshly-grown colonies were picked and used to inoculate 12 mL 2x TY+Amp+Cam media, incubated at 30 °C overnight. These starter cultures were then diluted 1:100 into large-scale 2x TY+Amp+Cam cultures, which were grown to OD₆₀₀ 0.8 at 37 °C, induced with 0.5 mM IPTG and incubated for a further 3 hr at 37 °C. Cells were then harvested by centrifugation at 7,500 g for 15 min at 4 °C and pellets stored at -80 °C until induction was confirmed on 18 % SDS-PAGE.

A2.2.3 Ni-NTA purification of soluble IF2-SA from BL21 (DE3) RIPL cells

The cell pellet from a 750 mL culture was thawed and resuspended in 10 mL 300 mM NaCl, 50 mM Tris, 0.1 mM phenylmethylsulfonyl fluoride (PMSF) in Complete Protease Inhibitor Cocktail (Roche, UK) and then sonicated for 3 min pulsed on ice at 40 % amplitude on a Sonics Vibra-Cell sonicator. Following centrifugation at 17,000 g for 15 min at 4 °C, the soluble fraction (supernatant) was incubated with Qiagen Ni-NTA Superflow resin (which was equilibrated in 300 mM NaCl, 50 mM Tris pH 7.8) for 1 hour at 4 °C. The resin was then isolated by centrifugation, washed with 300 mM NaCl, 50 mM Tris, 30 mM imidazole pH 7.8 and then added to a poly-prep column (Bio-Rad) for elution with 5 mL 300 mM NaCl, 50 mM Tris, 200 mM imidazole pH 7.8. The eluate was then dialysed three times against PBS and protein concentration determined from the absorbance at 280 nm (a 1 mg/mL solution of IF2-SA has a predicted A_{280} of 1.144 according to ProtParam (Gasteiger et al., 2005)). At each stage of the purification process, samples were taken for 8 % SDS-PAGE analysis.

A2.2.4 Enzymatic digestion of IF2-SA

Elastase (~ 26 kDa) and thermolysin (~ 35 kDa) are endopeptidases that can digest a wide variety of protein substrates. To test their ability at digesting IF2-SA to produce core-SA, 100 µL of 5 µM IF2-SA in 50 mM Tris pH 8 was digested in duplicate with elastase (isolated from porcine pancrease, supplied by Sigma, UK), thermolysin (isolated from *Bacillus thermoproteolyticus rokko*, supplied by Sigma, UK) or both enzymes, in a 1:50 w/w protease to protein ratio. Digests were incubated at the desired temperatures (37 °C for elastase and both 37 °C and 50 °C for thermolysin) for the indicated time periods (2

hours and overnight) and then boiled in SDS loading buffer and run on 18 % SDS-PAGE.

A2.2.5 Enzymatic digestion of soluble fraction of IF2-SA cell pellet

The cell pellet from a 750 mL IF2-SA culture was thawed and resuspended in 10 mL 50 mM Tris pH 8 and then sonicated for 3 min pulsed on ice at 40 % amplitude on a Sonics Vibra-Cell sonicator. Following centrifugation at 15,000 g for 15 min at 4 °C, the soluble fraction (supernatant) was incubated with 0.2 mg of elastase at 37 °C overnight.

A2.2.6 2-iminobiotin affinity chromatography of core-SA from IF2-SA digest

The pH of the elastase-digested soluble fraction of the IF2-SA cell pellet was adjusted to pH 11 using 1 M Na_2HPO_4 pH 11 buffer. 2 mL of a 50 % slurry of 2-iminobiotin-agarose (Sigma, UK), which was equilibrated in 50 mM Na_2HPO_4 , 0.5 M NaCl pH 11, was then added to the digest and incubated at 4 °C for 1 hour. The resin was then isolated by centrifugation, washed with 50 mM Na_2HPO_4 , 0.5 M NaCl pH 11 and then added to a poly-prep column (Bio-Rad, UK) for elution with 3 mL 50 mM sodium acetate pH 4. The eluate was then dialysed three times against PBS and protein concentration determined from the absorbance at 280 nm (a 1 mg/mL solution of streptavidin has A_{280} of 3.129 according to ProtParam (Gasteiger et al., 2005)).

A2.2.7 Site-directed mutagenesis to make IF2-Tr

The IF2-SA plasmid was used as the template DNA for site-directed mutagenesis, to introduce the Tr mutations S52G R53D into the SA sequence. Appropriate primers were designed to introduce the mutations into the SA sequence. All DNA oligonucleotides used in this study were synthesised by Invitrogen (Paisley, UK) and sequences given in Table 22. PCR was carried out in 25 μ L reactions in a Bio-Rad DNA Engine® Peltier Thermal Cycler (Hemel Hempstead, UK) using KOD DNA polymerase (EMD Chemicals, UK). A typical reaction would contain the following:

Purified IF2-SA template plasmid	5 ng
IF2-TR Forward primer	10 pmol
IF2-TR Reverse primer	10 pmol
10x KOD reaction buffer	2.5 μ L
MgSO ₄	1.5 μ L
10 mM dNTPs	2.5 μ L
KOD DNA polymerase	0.5 μ L
ddH ₂ O	to 25 μ L

Typical reaction conditions would be 95 °C for 3 minutes, followed by 20 cycles of 95 °C for 20 s, 64 °C for 30 s and 68 °C for 4 minutes. 5 μ L of the amplified reaction mixture underwent 0.7 % (w/v) agarose gel electrophoresis at 130 V in TAE buffer (40 mM Tris acetate, 1 mM EDTA pH 8.2) and PCR products were detected by ethidium bromide staining. 10 μ L of the amplified reaction mix was digested with 10 units of DpnI (New England Biolabs, UK) at 37 °C for 2 hours to digest methylated and hemimethylated DNA. 4 μ L of the digested mixture was then transformed into competent *E. coli* XL1-Blue cells according to the

manufacturer's protocol and transformants selected for on Luria Broth (LB) agar plates (1 % tryptone, 0.5 % yeast extract, 1 % NaCl, 1.6 % agar) supplemented with 100 µg/mL ampicillin (Amp). Individual colonies were picked and used to inoculate 12 mL of LB containing 100 µg/mL Amp, which was cultured overnight at 37 °C in a shaking incubator. Cells were then harvested by centrifugation at 4,000 g for 5 minutes and DNA extracted using a Qiagen Miniprep kit (Valencia, CA). Sequences were confirmed by sequencing with T7 forward and reverse primers (Table 24) by Geneservice Ltd., UK.

A2.2.8 IF2-Tr expression trials

Small-scale expression trials were conducted in order to identify the optimal conditions for soluble production of the correctly-folded IF2-Tr tetramer. The expression conditions tested are given in Table 24. Briefly, the IF2-Tr plasmid was subcloned into two *E. coli* expression strains according to the manufacturer's protocol: BL21 (DE3) RIPL cells and B834 (DE3) pLysS cells (see Table 25 for details). Transformants were selected on either 2x TY agar plates (1.6 % tryptone, 1 % yeast extract, 0.5 % NaCl, 1.5 % agar) or LB agar plates (1 % tryptone, 0.5 % yeast extract, 1 % NaCl, 1.6 % agar) (depending on the culture medium being tested), both supplemented with 100 µg/mL Amp and 25 µg/mL Cam. Individual freshly-grown colonies were picked and used to inoculate 12 mL of the appropriate media (either 2x TY+Amp+Cam or LB+Amp+Cam) and incubated at 30 °C overnight. These starter cultures were then diluted 1:100 into 200 mL small-scale expression cultures, in the appropriate media. Cultures were grown at 37 °C until the induction O.D. was reached, when expression was induced according to the conditions in Table 26.

After induction, cells were then harvested by centrifugation at 7,500 g for 15 min at 4 °C and pellets stored at -80 °C until induction was confirmed on 6 % SDS-PAGE (unboiled samples).

A2.2.9 Optimised IF2-Tr expression and purification

Following expression trials, the IF2-Tr plasmid was subcloned into B834 (DE3) pLysS cells according to the manufacturer's protocol. Transformants were selected for on 2x TY agar plates supplemented with 100 µg/mL Amp and 25 µg/mL Cam. Individual freshly-grown colonies were picked and used to inoculate 12 mL 2x TY+Amp+Cam media containing 0.8 % glucose, incubated at 30 °C overnight. These starter cultures were then diluted 1:100 into large-scale 2x TY+Amp+Cam cultures, which were grown to OD₆₀₀ 0.9 at 37 °C, induced with 0.5 mM IPTG and incubated for a further 3 hr at 37 °C. Cells were then harvested by centrifugation at 7,500 g for 15 min at 4 °C and pellets stored at -80 °C until induction was confirmed on 18 % SDS-PAGE.

A2.2.10 Ni-NTA purification of soluble IF2-Tr from B834 (DE3) pLysS cells

The cell pellet from a 750 mL culture was thawed and resuspended in 10 mL 300 mM NaCl, 50 mM Tris, 0.1 mM phenylmethylsulfonyl fluoride (PMSF) in Complete Protease Inhibitor Cocktail (Roche, UK) and then sonicated for 3 min pulsed on ice at 40 % amplitude on a Sonics Vibra-Cell sonicator. Following centrifugation at 17,000 g for 15 min at 4 °C, the soluble fraction (supernatant) was incubated with Qiagen Ni-NTA Superflow resin (which was equilibrated in 300 mM NaCl, 50 mM Tris pH 7.8) overnight at 4 °C. The resin was then isolated by centrifugation, washed with 300 mM NaCl, 50 mM Tris, 10 mM

imidazole pH 7.8 and then added to a poly-prep column (Bio-Rad) for elution with 5 mL 300 mM NaCl, 50 mM Tris, 200 mM imidazole pH 7.8. The eluate was then dialysed three times against PBS and protein concentration determined from the absorbance at 280 nm (a 1 mg/mL solution of IF2-SA has a predicted A_{280} of 1.144 according to ProtParam (Gasteiger et al., 2005)). At each stage of the purification process, samples were taken for 6 % SDS-PAGE analysis.

A2.2.11 Enzymatic digestion of soluble fraction of IF2-Tr cell pellet

The cell pellet from a 750 mL IF2-Tr culture was thawed and resuspended in 10 mL 50 mM Tris pH 8 and then sonicated for 3 min pulsed on ice at 40 % amplitude on a Sonics Vibra-Cell sonicator. Following centrifugation at 15,000 g for 15 min at 4 °C, the soluble fraction (supernatant) was incubated with 0.2 mg of elastase at 37 °C overnight. At each stage, samples were taken for 8 % (unboiled samples) and 18 % (boiled samples) SDS-PAGE analysis.

Name	Sequence (5' – 3')
T7 F	TAATACGACTCACTATAGGG
T7 R	GCTAGTTATTGCTCAGCGG
IF2-TR F	GGCAACGCCGAGGGTGACTACGTCCTGACCGGTCG
IF2-TR R	CGACCGGTCAGGACGTAGTCACCCTCGGCGTTGCC

Table 24. DNA oligonucleotides used in this study.

Bacterial strain	Description	Antibiotic resistance
BL21 (DE3) RIPL	General purpose expression host	Chloramphenicol ^R
BL21 (DE3) pLysS	High-stringency expression host	Chloramphenicol ^R
B834 (DE3) pLysS	Protease-deficient expression host	Chloramphenicol ^R

Table 25. Bacterial strains used in this study.

Culture ID	Bacterial strain	Media	Induction conditions:			
			O.D.	[IPTG] (mM)	Temp (°C)	Time (hr)

<i>First round of IF2-Tr expression trials</i>						
A	BL21 (DE3) RIPL	LB	0.5	0.5	30	3
B	B834 (DE3) pLysS	LB	0.5	0.5	30	3
C	BL21 (DE3) RIPL	LB	0.5	0.2	30	O/N
D	BL21 (DE3) RIPL	LB	0.5	0.05	30	O/N
E	BL21 (DE3) RIPL	LB	0.5	0.5	30	6
F	BL21 (DE3) RIPL	LB	0.5	0.5	30	O/N
G	BL21 (DE3) RIPL	2TY	0.5	0.5	30	3
H	BL21 (DE3) RIPL	LB	0.7	0.5	30	3
I	BL21 (DE3) RIPL	LB	0.9	0.5	30	3
J	BL21 (DE3) RIPL	LB	0.5	0.5	25	O/N
K	BL21 (DE3) RIPL	LB+10 mM Imid	0.5	0.5	30	3
L	BL21 (DE3) RIPL	LB+1 mM Imid	0.5	0.5	30	3
<i>Second round of IF2-Tr expression trials</i>						
A	B834 (DE3) pLysS	LB	0.5	0.5	30	3
B	B834 (DE3) pLysS	2TY	0.9	0.5	30	3
C	B834 (DE3) pLysS	2TY	1.3	0.5	30	3
D	B834 (DE3) pLysS	2TY	0.9	1.0	30	3
E	B834 (DE3) pLysS	2TY	0.9	0.5	37	3

Table 26. IF2-Tr expression conditions investigated. All cultures were grown with ampicillin and chloramphenicol at 37 °C to induction O.D.

A2.3 Results

A2.3.1 Expression of IF2-SA

The plasmid encoding domain I of IF2 fused to the N-terminal of SA (hereafter called IF2-SA) was transformed into three *E. coli* strains for small-scale expression trials; BL21 (DE3) RIPL, B834 (DE3) pLysS and BL21 (DE3) pLysS. The cultures were grown at 37 °C until an O.D. of 0.8, followed by induction for 3 hours at 37 °C with 0.5 mM IPTG. These conditions are considered the least favourable conditions for soluble expression. All three bacterial strains gave strong induction (Figure 77). IF2-SA was then purified from BL21 (DE3) RIPL cells.

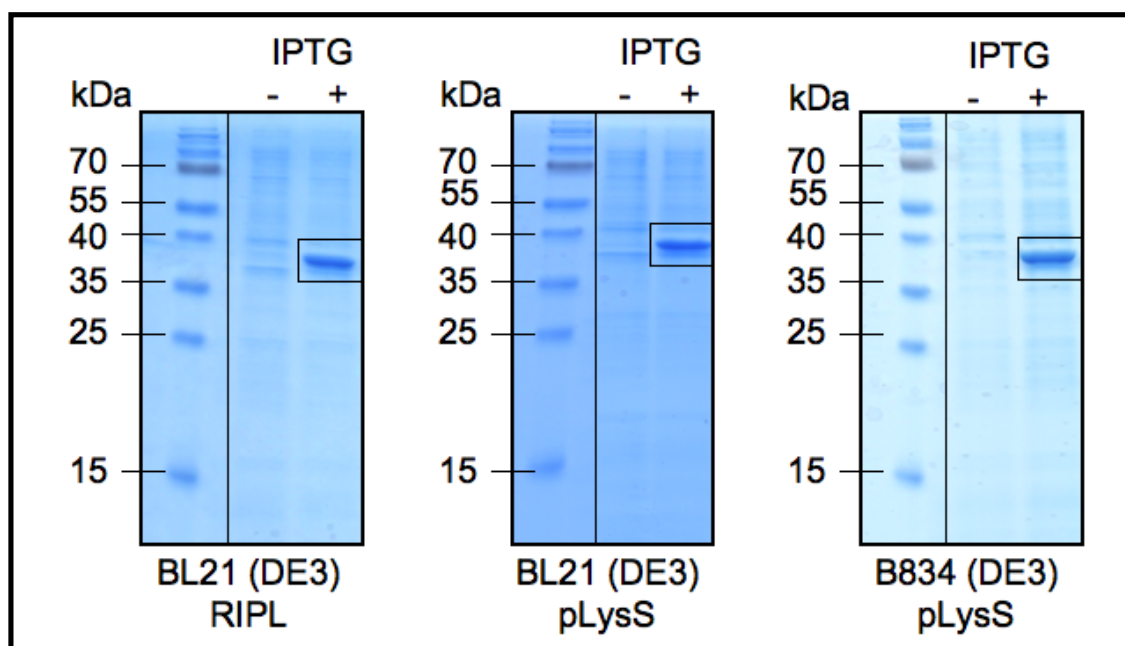


Figure 77. A three-hour induction with IPTG at 37 °C gave strong induction of IF2-SA in BL21 (DE3) RIPL cells (left), BL21 (DE3) pLysS cells (middle) and B834 (DE3) pLysS cells (right). Samples of IF2-SA expression cultures before and after IPTG induction were boiled with 6x SDS loading buffer and run on 18 % SDS-PAGE. The band corresponding to monomeric IF2-SA (due to boiling before SDS-PAGE) is highlighted.

A2.3.2 Purification of IF2-SA from BL21 (DE3) RIPL cells

Although far too much protein was loaded onto the SDS-PAGE gel, it appears that at least half of the total protein expressed in BL21 (DE3) RIPL cells is soluble (Figure 78).

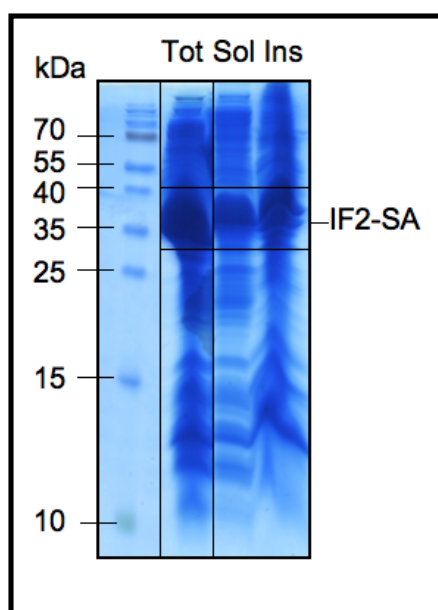


Figure 78. A substantial fraction of the total IF2-SA protein expressed in BL21 (DE3) RIPL cells is soluble. Expression was induced with IPTG for 3 hours at 37 °C. After cell lysis by sonication and centrifugation, the soluble (sol) and insoluble (ins) fractions were boiled with SDS loading buffer and then run on 18 % SDS-PAGE, to compare to the total protein present (tot).

Ni-NTA purification was then carried out on the soluble IF2-SA protein (Figure 79), revealing that the Ni-NTA capture of IF2-SA was highly inefficient. Also, quite a large amount of IF2-SA was lost on washing with 30 mM imidazole-containing buffer. On elution, the purified IF2-SA gave multiple bands on the gel, suggesting the washing steps were not sufficient to remove all contaminating proteins from the sample.

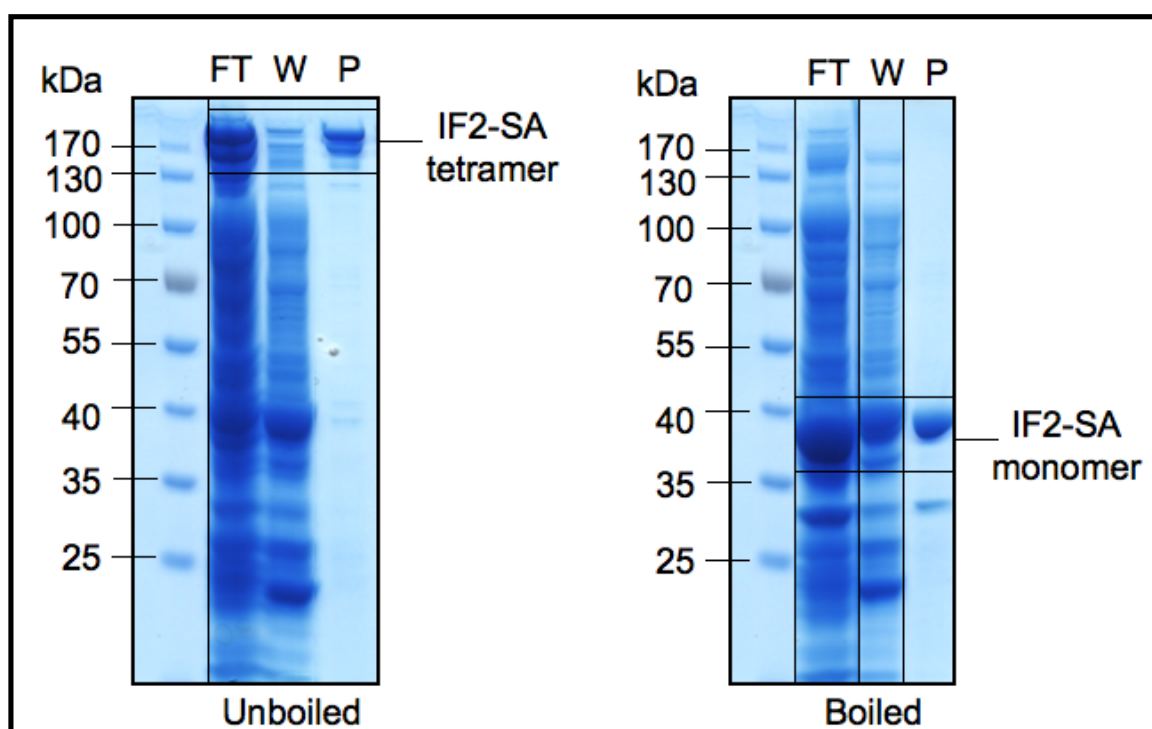


Figure 79. Ni-NTA purification of soluble IF2-SA from BL21 (DE3) RIPL cells resulted in a large amount of IF2-SA being present in the flow-through (FT) and 30 mM imidazole wash (W). 18 % SDS-PAGE with unboiled samples (left) and samples boiled in SDS loading buffer prior to loading (right). Bands corresponding to IF2-SA are highlighted. In the unboiled samples (left), IF2-SA does not migrate according to its molecular weight, in agreement with earlier observations (Sorensen et al., 2003). The final purified protein (P) gave multiple bands on the gel, indicating more stringent washing steps are needed in the purification protocol.

The final yield of soluble IF2-SA from BL21 (DE3) RIPL cells using Ni-NTA purification was 10 mg/L.

A2.3.3 Enzymatic digestion of IF2-SA

The endopeptidases elastase and thermolysin were tested for their ability to digest purified IF2-SA. Due to the resistance of SA to proteolytic digestion, it was hypothesised that non-specific proteolytic digestion of IF2-SA by elastase or thermolysin would liberate core-SA. A two-hour incubation of IF2-SA at 37 °C with the enzymes, individually and in combination, revealed this hypothesis was correct, with core-SA being produced (Figure 80).

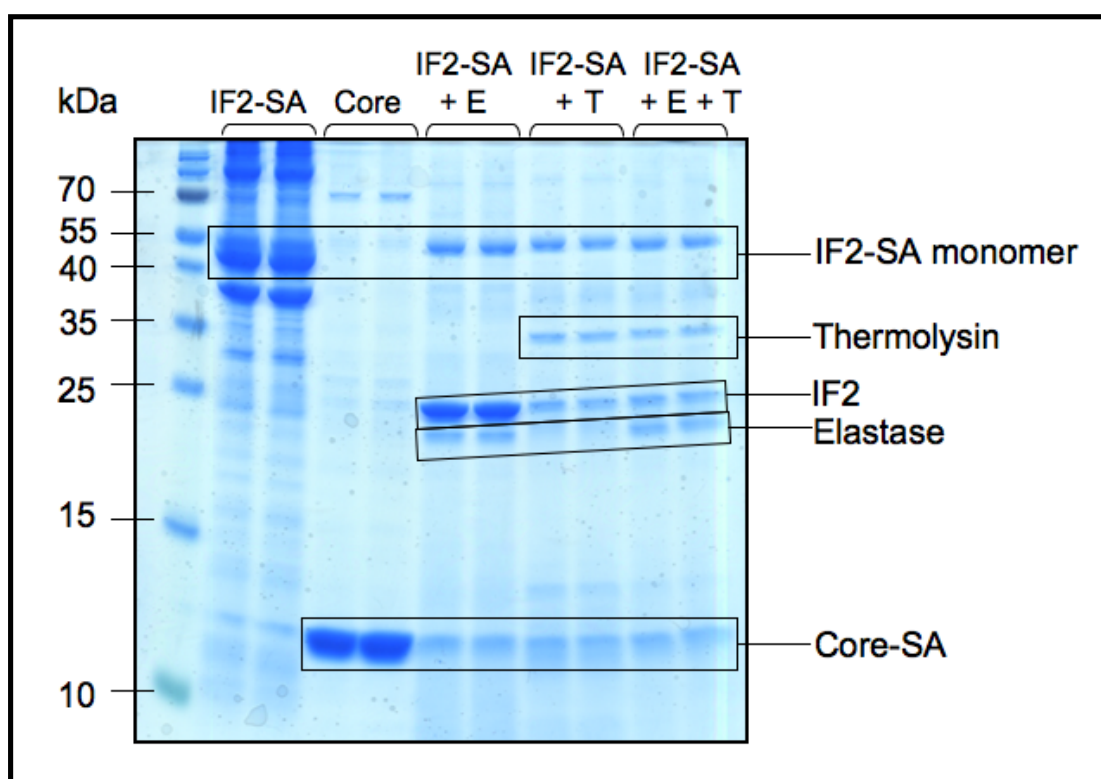


Figure 80. Enzymatic digestion of IF2-SA with elastase (E) and thermolysin (T) generates core-SA. Enzymes were incubated with IF2-SA for 2 hours at 37 °C. All samples were boiled with SDS loading buffer before 18 % SDS-PAGE. For comparison, core-SA purified from inclusion bodies was also boiled with SDS loading buffer and run on 18 % SDS-PAGE.

The effect of increasing the digestion period to overnight at 37 °C and increasing the incubation temperature for thermolysin digestion to 50 °C was then investigated (Figure 81). This revealed that overnight digestion gave a stronger band corresponding to core-SA than two-hour digestion (Figure 80), suggesting that a longer incubation period could increase the yield of core-SA from IF2-SA digestion. The core-SA bands from overnight digestion with thermolysin at 50 °C appeared slightly weaker than the 37 °C digests. However, digestion with elastase, thermolysin and both enzymes at 37 °C all appeared to give equally strong bands for core-SA on SDS-PAGE. Therefore, it was decided that overnight digestion at 37 °C with elastase would be used for future purifications.

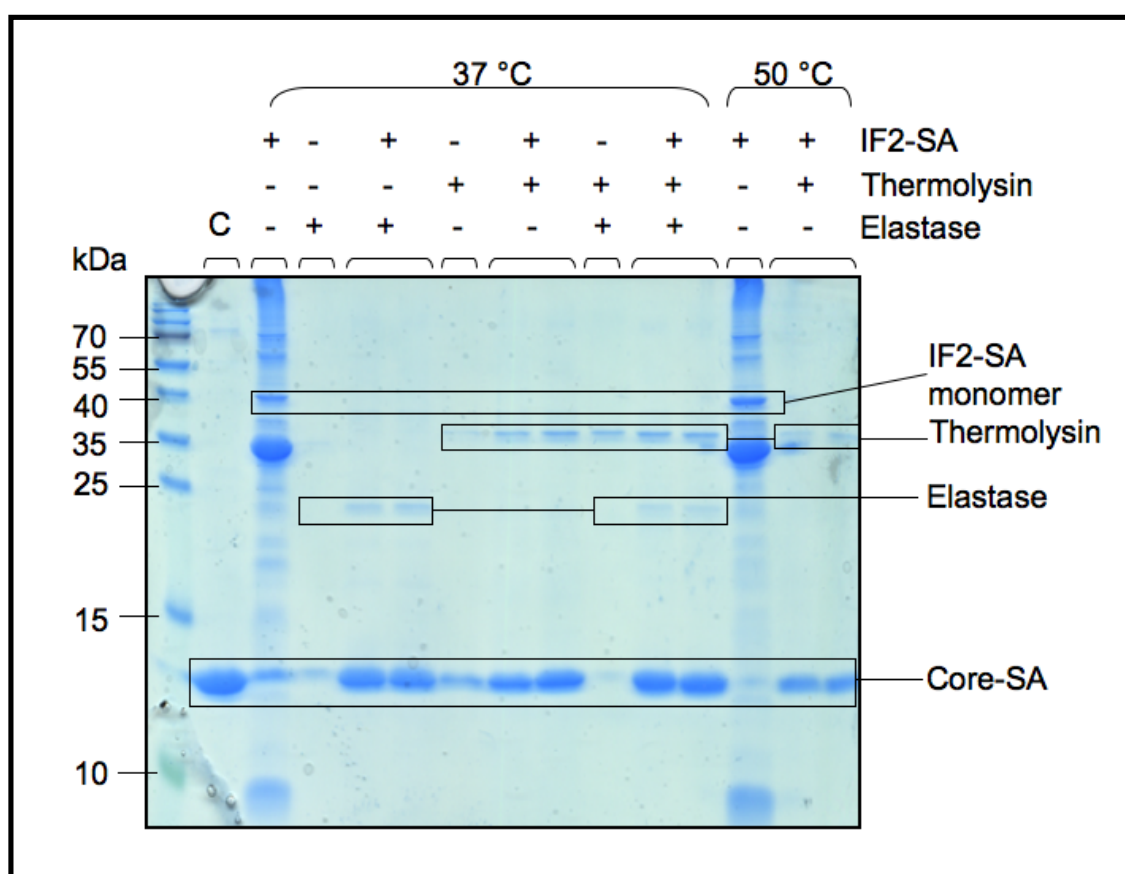


Figure 81. Overnight enzymatic digestion of IF2-SA with elastase (E) at 37 °C and thermolysin (T) at 50 °C generates core-SA. All samples were boiled with SDS loading buffer before 18 % SDS-PAGE. The control lane (C) contained only core-SA purified from inclusion bodies.

The IF2-SA sample used for digestion testing was extremely impure, and it was striking how efficiently these contaminating proteins were degraded by enzymatic digestion (Figure 81). Therefore, I investigated whether enzymatic digestion of the unpurified soluble fraction of an IF2-SA cell pellet would degrade all contaminating proteins in the soluble fraction, as well as the IF2-tag and linkers, leaving only core-SA. This would be advantageous, as it would avoid the use of Ni-NTA purification of IF2-SA prior to digestion, which was earlier shown to be inefficient at capturing IF2-SA protein (Figure 79). After elastase digestion of the soluble fraction at 37 °C overnight, 2-iminobiotin affinity chromatography was then carried out to purify core-SA (Figure 82). All core-SA from the digest was captured by the 2-iminobiotin column (at pH 11), and was retained on the column during the washing steps, eluting only when the pH was lowered to pH 4. A small amount of IF2 monomer also eluted from the column. In order for this IF2 to have been retained on the column it was most likely tagged to SA (i.e. as IF2-SA) during the capture step, becoming digested by carried-over elastase during the washing steps.

The final yield of core-SA from elastase digestion of the soluble fraction of a IF2-SA cell pellet was 1 mg/L.

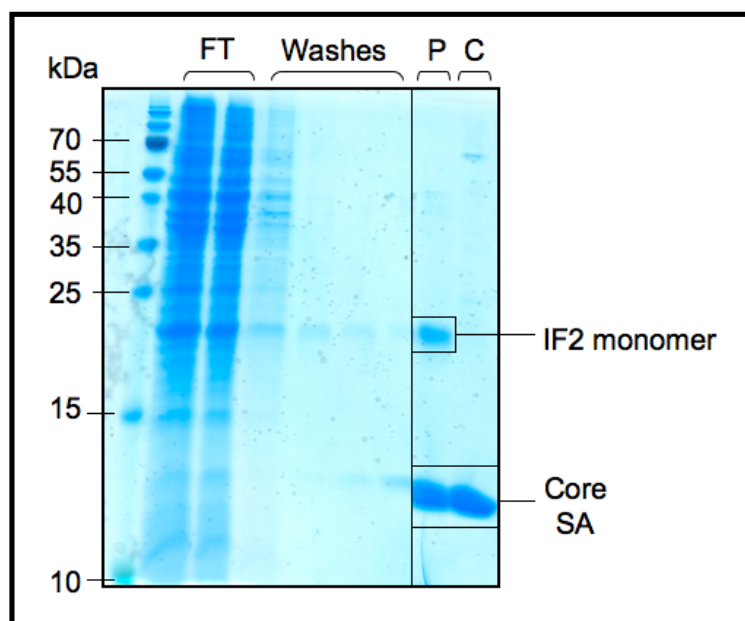


Figure 82. 2-iminobiotin affinity chromatography on elastase-digested soluble fraction of IF2-SA cell pellet. All core-SA in the digest was captured by the resin at pH 11, with none lost on washing. Core-SA (P) was eluted from the column at pH 4. The control lane (C) contained only core-SA purified from inclusion bodies.

A2.3.4 Putting Tr mutations into IF2-SA plasmid

The IF2-SA plasmid was used as the template for site-directed mutagenesis to introduce the Tr mutations S52G R53D into the SA sequence. Sequencing confirmed the successful production of the plasmid encoding domain I of IF2 fused to the N-terminal of Tr (hereafter called IF2-Tr).

A2.3.5 IF2-Tr expression trials

The inefficient Ni-NTA capture of IF2-SA suggested that the His₆ tag was inaccessible to the column resin (see discussion). It was possible that IF2-SA was misfolding, which could result in the His₆ tag being obscured. Therefore, it was decided to conduct expression trials of IF2-Tr to identify expression conditions that give optimum IF2-Tr tetramer formation.

200 mL IF2-Tr cultures were grown under a number of conditions in the first round of expression trials (Table 27). After IPTG induction, samples were taken and run on SDS-PAGE (Figure 83). The conditions tested showed a wide variety of expression levels, from no IF2-Tr tetramer production to strong IF2-Tr tetramer production. For example, under the same conditions, B834 (DE3) pLysS cells gave a larger amount of folded IF2-Tr tetramer than BL21 (DE3) RIPL cells.

Culture ID	Bacterial strain	Media	Induction conditions:			
			O.D.	[IPTG] (mM)	Temp (°C)	Time (hr)
A	BL21 (DE3) RIPL	LB	0.5	0.5	30	3
B	B834 (DE3) pLysS	LB	0.5	0.5	30	3
C	BL21 (DE3) RIPL	LB	0.5	0.2	30	O/N
D	BL21 (DE3) RIPL	LB	0.5	0.05	30	O/N
E	BL21 (DE3) RIPL	LB	0.5	0.5	30	6
F	BL21 (DE3) RIPL	LB	0.5	0.5	30	O/N
G	BL21 (DE3) RIPL	2TY	0.5	0.5	30	3
H	BL21 (DE3) RIPL	LB	0.7	0.5	30	3
I	BL21 (DE3) RIPL	LB	0.9	0.5	30	3
J	BL21 (DE3) RIPL	LB	0.5	0.5	25	O/N
K	BL21 (DE3) RIPL	LB+10 mM Imid	0.5	0.5	30	3
L	BL21 (DE3) RIPL	LB+1 mM Imid	0.5	0.5	30	3

Table 27. A variety of conditions were tested in the first round of IF2-Tr expression trials. All cultures were grown with ampicillin and chloramphenicol at 37 °C to induction O.D.

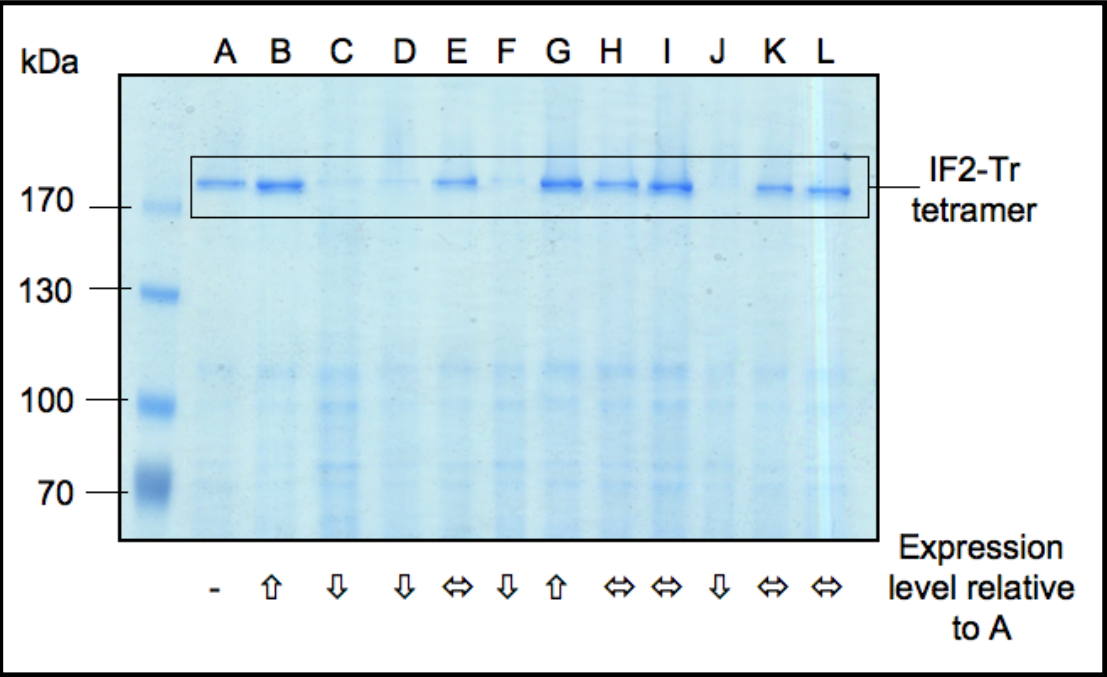


Figure 83. The conditions tested in the first round of expression trials gave a variety of levels of tetrameric IF2-Tr production. Post-induction samples were mixed with SDS loading buffer and run on 6 % SDS-PAGE. The expression levels relative to condition A are represented by arrows: an up arrow represents higher IF2-Tr tetramer production, a double-headed horizontal arrow represents a IF2-Tr tetramer level equivalent to A, and a down arrow represents a lower level of tetrameric IF2-Tr production. Conditions A-L are as detailed in Table 27.

A second round of expression trials was then conducted, testing a number of conditions, in duplicate, using B834 (DE3) pLysS cells (Table 28). As in earlier expression trials, post-induction samples were taken and analysed by SDS-PAGE (Figure 84). Growth in 2TY with induction at 30 °C at O.D. 0.9 (condition B) and with double the concentration of IPTG used previously (condition D) were shown to give the highest levels of IF2-Tr tetramer formation.

Culture ID	Media	O.D.	Induction conditions:		
			[IPTG] (mM)	Temp (°C)	Time (hr)
A	LB	0.5	0.5	30	3
B	2TY	0.9	0.5	30	3
C	2TY	1.3	0.5	30	3
D	2TY	0.9	1.0	30	3
E	2TY	0.9	0.5	37	3

Table 28. Five expression conditions were tested in the second round of IF2-Tr expression trials. All B834 (DE3) pLysS cultures were grown with ampicillin and chloramphenicol at 37 °C to induction O.D.

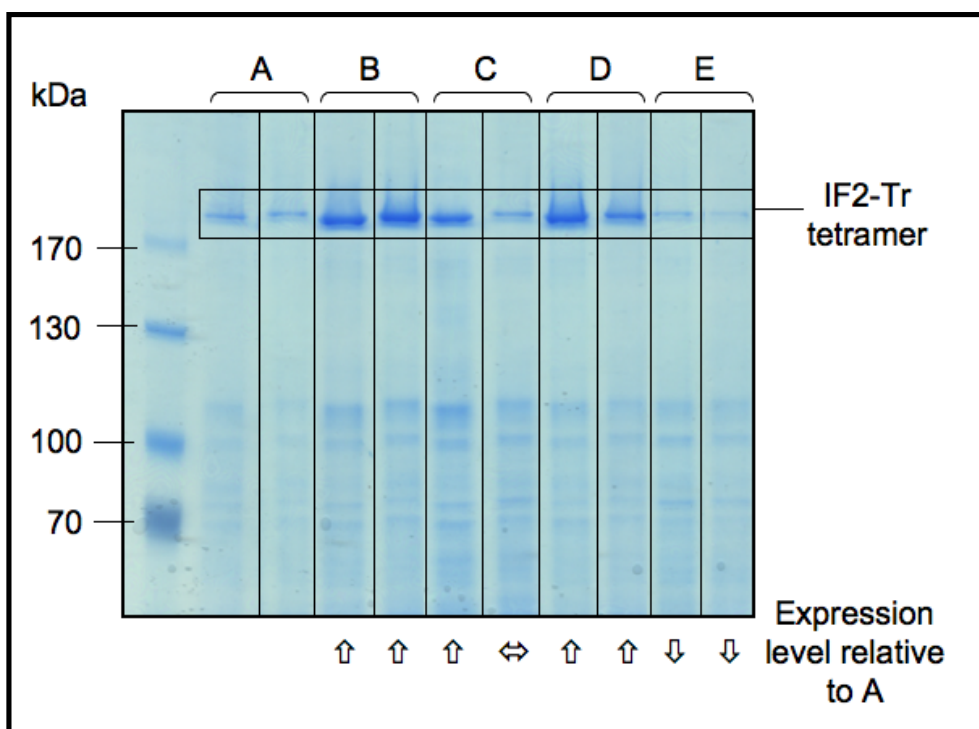


Figure 84. A smaller selection of conditions were tested, in duplicate, in the second round of IF2-Tr expression trials. Post-induction samples were mixed with SDS loading buffer and run on 6 % SDS-PAGE. The expression levels relative to condition A are represented by arrows: an up arrow represents higher IF2-Tr tetramer production, a double-headed horizontal arrow represents a IF2-Tr tetramer level equivalent to A, and a down arrow represents a lower level of tetrameric IF2-Tr production. Conditions A-E are as detailed in Table 28.

As a result of the expression trials, it was decided that condition B would be used for IF2-Tr large-scale expression.

A2.3.6 IF2-Tr expression and purification

Using expression condition B detailed above, IF2-Tr was expressed in B834 (DE3) pLysS cells at 30 °C for 3 hours. After induction with 0.5 mM IPTG, samples were taken and run on SDS-PAGE, with and without boiling (Figure 85). The non-boiled sample revealed that there was some correctly folded IF2-Tr post-induction, but the majority of IF2-Tr protein was mis-folded.

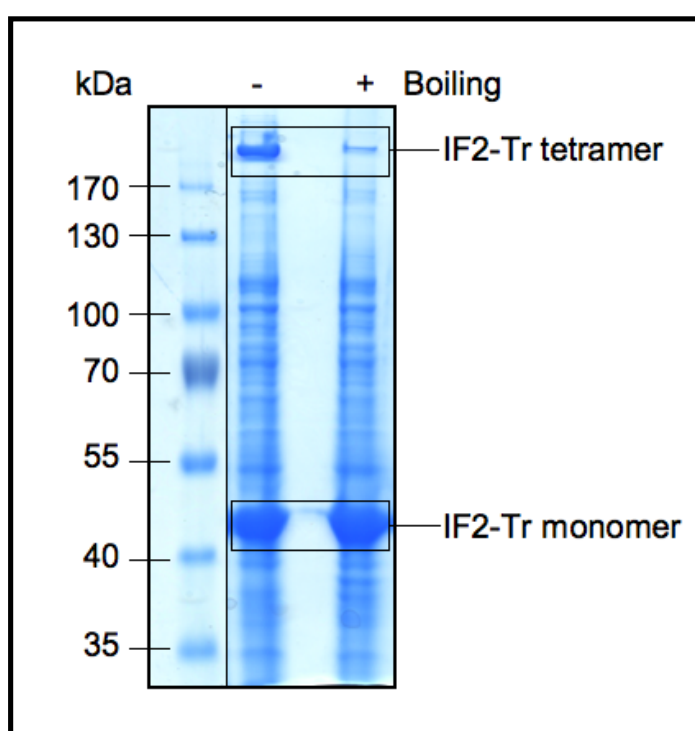


Figure 85. Expression cultures of IF2-Tr were analysed by SDS-PAGE with and without boiling prior to loading, to assess the level of IF2-Tr misfolding during expression. Without boiling the sample prior to SDS-PAGE, a large amount of IF2-Tr protein is misfolded and unable to form a tetramer, with the band corresponding to IF2-Tr monomer being considerably larger than the band corresponding to IF2-Tr tetramer.

Ni-NTA purification was attempted using the soluble fraction of the IF2-Tr cell pellet (Figure 86). As with IF2-SA, there was not efficient capture of IF2-Tr on the resin, with some IF2-Tr present in the column flow-through and also in the early washing steps. This was to be expected due to the large amounts of misfolded IF2-Tr monomer in the soluble fraction. The final eluate from the column contained IF2-Tr but the protein was very impure, with many contaminating proteins also present. This was most likely due to the washing buffer containing a lower concentration of imidazole than normal (10 mM rather than the standard 30 mM imidazole), in an effort to reduce the amount of captured IF2-Tr being lost in the washing step.

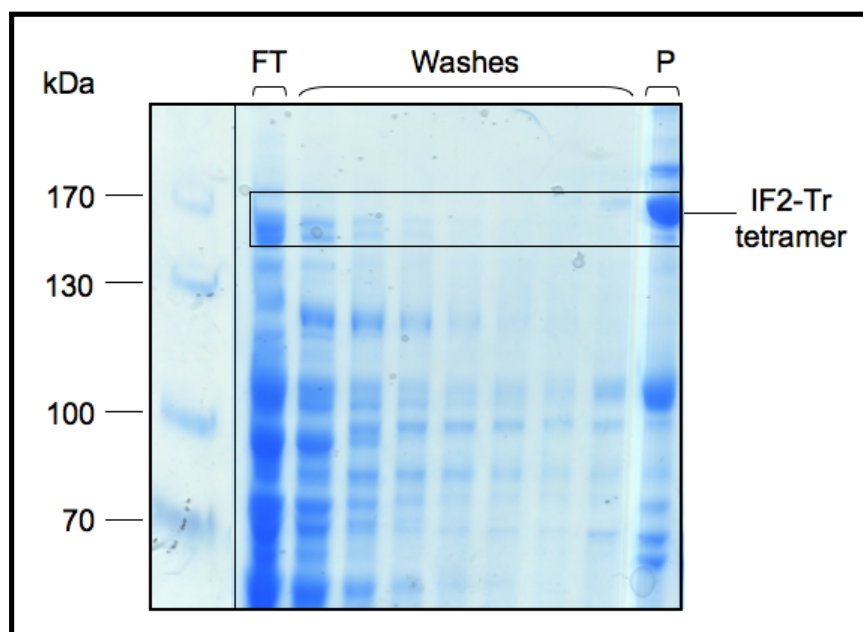


Figure 86. Ni-NTA purification of the IF2-Tr protein from the soluble fraction of B834 (DE3) pLysS cells was not successful. Not all IF2-Tr present was captured on the Ni-NTA, as some protein was present in the column flow-through (FT). Extensive washes with 10 mM imidazole-containing wash buffer did remove some contaminating proteins, but also caused some captured IF2-Tr to be lost. The column eluate (P) gave an impure IF2-Tr sample. As with IF2-SA, IF2-Tr does not migrate according to its molecular weight. All samples were mixed with SDS loading buffer prior to 6 % SDS-PAGE.

A2.3.7 Enzymatic digestion of IF2-Tr

As Ni-NTA affinity chromatography failed to purify IF2-Tr, I attempted enzymatic digestion of the unpurified soluble fraction of a IF2-Tr cell pellet using elastase. After incubating the soluble cell fraction with elastase overnight at 37 °C, 2-iminobiotin affinity chromatography was used to purify the core-Tr produced (Figure 87). This revealed all core-Tr from the digest was captured by the 2-iminobiotin column (at pH 11), and was retained on the column during the washing steps, eluting only when the pH was lowered to pH 4.

The final yield of core-Tr from elastase digestion of the soluble fraction of a IF2-Tr cell pellet was 1 mg/L.

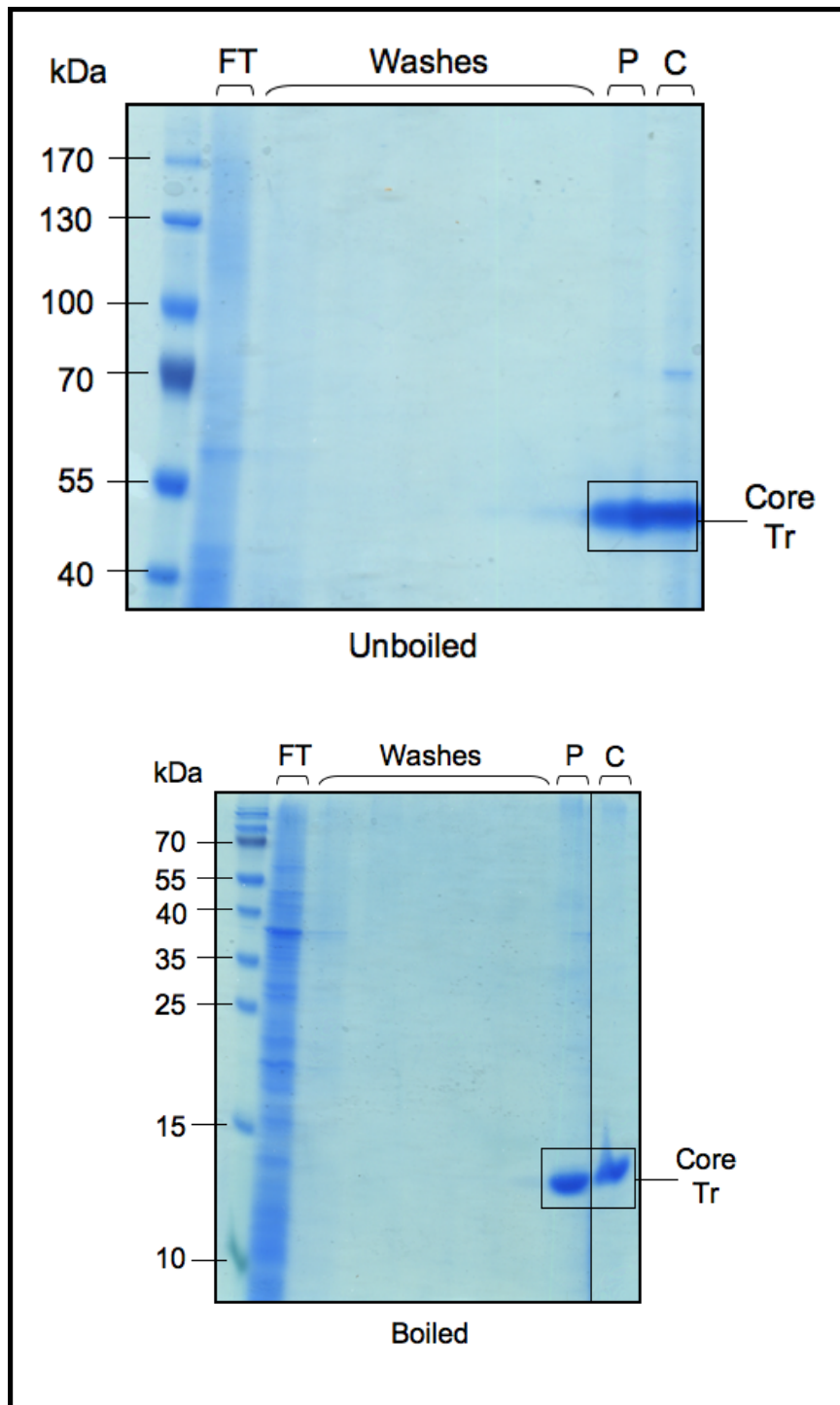


Figure 87. 2-iminobiotin affinity chromatography on elastase-digested soluble fraction of IF2-Tr cell pellet. All core-Tr in the digest was captured by the resin at pH 11, with none lost on washing. Core-Tr (P) was eluted from the column at pH 4. The control lane (C) contained only core-Tr purified from inclusion bodies.

A2.4 Discussion

On testing the recombinant expression of IF2-SA, over half of the total protein was expressed solubly in BL21 (DE3) RIPL cells. This is in agreement with, if not better than, the 47 % soluble IF2-SA expression level reported previously using the same expression conditions (Sorensen et al., 2003).

However, the trial purification carried out on IF2-SA revealed a difficulty in capturing the protein on Ni-NTA resin. Protein samples were rotated with Ni-NTA at 4 °C for one hour, and then overnight, but in both cases a large amount of IF2-SA was not captured by the resin. As this suggested possible protein misfolding (which would obscure the His₆ tag), expression trials of the IF2-Tr protein were conducted prior to purification attempts, in the hope of identifying conditions that would give minimal misfolding. However, the conditions highlighted by the expression trials still gave a large amount of misfolded IF2-Tr monomer, which was inefficiently captured on Ni-NTA resin. Further work is needed on reducing this misfolding, in order to increase the protein yield. The protein construct could be altered, e.g. increasing the linker length between the IF2 domain and SA / Tr, in case steric hindrance from the IF2 domain is preventing tetramer formation. Other solubility fusion partners could also be investigated, such as MBP (Kapust and Waugh, 1999) and NusA (Davis et al., 1999), as although these were previously shown to produce inactive soluble SA (Sorensen et al., 2003), they are commonly used for soluble expression of many other proteins. Other solubility fusion partners that could be tested include thioredoxin (LaVallie et al., 2000) and protein disulfide isomerase I (Zhang et al., 1998).

The induction conditions used in this work (0.5 mM IPTG at 30 °C for 3 hours) were certainly not favourable for soluble expression. Transcription and translation are tightly coupled in *E. coli*, and it has been calculated that the minimal rate of protein synthesis in an average *E. coli* cell is ~ 60,000 nascent polypeptide chains each minute (Lorimer, 1996). If a strong expression promoter is used together with a high inducer concentration, the rapid rate of protein synthesis means that the high concentration of newly synthesised protein is at risk of aggregating before it has the chance to fold. This could be occurring in the IF2-SA / IF2-Tr expressions tested here. Reducing the rate of protein synthesis will result in the newly produced protein having more time to fold and increases the likelihood of the protein remaining soluble. This can be achieved by: lowering the induction temperature (which slows all cellular processes including protein synthesis), with a corresponding increase in induction time (Francis and Page, 2010); and decreasing the concentration of inducer, which will slow the rate of transcription (Turner et al., 2005). Also, to facilitate correct folding, molecular chaperones such as GroEL/ES could be utilised, e.g. by using an expression vector that also encodes for certain *E. coli* chaperones (Kyratsous et al., 2009) or by using host cells that express chaperonins, such as Stratagene's ArcticExpress cells, which co-express the cold-adapted chaperonins Cpn10 and Cpn60 for improved protein folding at low temperatures. There are certainly more avenues to explore in the search for an optimal method of soluble SA expression.

However, a large amount of soluble IF2-SA in the bacterial cell is most likely to be toxic to the cell, due to SA binding the biotin carboxyl carrier protein (BCCP)

(Fall et al., 1971) and preventing its essential function (Wang et al., 2005). To avoid this, IF2-SA would have to be rapidly exported from the cell once produced and folded. It is therefore anticipated that, in addition to optimising soluble SA expression, the secretion strategy will have to be investigated (Mergulhao et al., 2005), for example, by investigating the coexpression of the *kil* gene, encoding for a bacteriocin release protein, which has previously been shown to increase the secretion of SA from *E. coli* (Miksch et al., 2008).

Despite encountering difficulties in expressing correctly-folded soluble SA and Tr, this work did demonstrate the utility of directly digesting the soluble fraction of a cell pellet with elastase prior to any purification steps, in order to degrade contaminating proteins. Subsequent purification steps, such as the 2-iminobiotin affinity column used in this work, can then be used to isolate the protein of interest directly from the digested soluble fraction. Using optimal expression conditions, this protocol should rapidly and easily give large yields of core-Tr. This will increase the utility of Tr, as a facile method of production should increase the attractiveness of Tr to researchers and companies alike.

References

- (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **50**, 760-763.
- Adams, P.D., Grosse-Kunstleve, R.W., Hung, L.W., Ioerger, T.R., McCoy, A.J., Moriarty, N.W., Read, R.J., Sacchettini, J.C., Sauter, N.K., and Terwilliger, T.C. (2002a). PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr* **58**, 1948-1954.
- Adams, S.R., Campbell, R.E., Gross, L.A., Martin, B.R., Walkup, G.K., Yao, Y., Llopis, J., and Tsien, R.Y. (2002b). New biarsenical ligands and tetracysteine motifs for protein labeling in vitro and in vivo: synthesis and biological applications. *J Am Chem Soc* **124**, 6063-6076.
- Adams, T.E., Epa, V.C., Garrett, T.P., and Ward, C.W. (2000). Structure and function of the type 1 insulin-like growth factor receptor. *Cell Mol Life Sci* **57**, 1050-1093.
- Ahlroth, M.K., Ahlroth, P., and Kulomaa, M.S. (2001). Copy-number fluctuation by unequal crossing-over in the chicken avidin gene family. *Biochem Biophys Res Commun* **288**, 400-406.
- Alcaraz, J., Buscemi, L., Puig-de-Morales, M., Colchero, J., Baro, A., and Navajas, D. (2002). Correction of microrheological measurements of soft samples with atomic force microscopy for the hydrodynamic drag on the cantilever. *Langmuir* **18**, 716-721.
- Almagro, J.C., and Fransson, J. (2008). Humanization of antibodies. *Front Biosci* **13**, 1619-1633.
- Alsop, E., Silver, M., and Livesay, D.R. (2003). Optimized electrostatic surfaces parallel increased thermostability: a structural bioinformatic analysis. *Protein Eng* **16**, 871-874.
- Andresen, M., Schmitz-Salue, R., and Jakobs, S. (2004). Short tetracysteine tags to beta-tubulin demonstrate the significance of small labels for live cell imaging. *Mol Biol Cell* **15**, 5616-5622.
- Argarana, C.E., Kuntz, I.D., Birken, S., Axel, R., and Cantor, C.R. (1986). Molecular cloning and nucleotide sequence of the streptavidin gene. *Nucleic Acids Res* **14**, 1871-1882.
- Arnold, F.H. (1998). When blind is better: protein design by evolution. *Nat Biotechnol* **16**, 617-618.
- Aslan, F.M., Yu, Y., Mohr, S.C., and Cantor, C.R. (2005). Engineered single-chain dimeric streptavidins with an unexpected strong preference for biotin-4-fluorescein. *Proc Natl Acad Sci U S A* **102**, 8507-8512.

- Aussel, L., Barre, F.X., Aroyo, M., Stasiak, A., Stasiak, A.Z., and Sherratt, D. (2002). FtsK Is a DNA motor protein that activates chromosome dimer resolution by switching the catalytic state of the XerC and XerD recombinases. *Cell* **108**, 195-205.
- Avrantinis, S.K., Stafford, R.L., Tian, X., and Weiss, G.A. (2002). Dissecting the streptavidin-biotin interaction by phage-displayed shotgun scanning. *Chembiochem* **3**, 1229-1234.
- Baggaley, K.H., Blessington, B., Flashaw, C.P., Ollis, W.D., Chalet, L., and Wolf, F.J. (1969). The constitution of stravidin, a novel microbiological product. *J Chem Soc D* **3**, 101-102.
- Bain, J.D., Glabe, C.G., Dix, T.A., and Chamberlin, A.R. (1989). Biosynthetic Site-Specific Incorporation of a Non-Natural Amino Acid into a Polypeptide. *J Am Chem Soc* **111**, 8013-8014.
- Barker, D.F., and Campbell, A.M. (1981). The birA gene of *Escherichia coli* encodes a biotin holoenzyme synthetase. *J Mol Biol* **146**, 451-467.
- Barroso, M.M. (2011). Quantum dots in cell biology. *J Histochem Cytochem* **59**, 237-251.
- Bayer, E.A., Ben-Hur, H., Hiller, Y., and Wilchek, M. (1989). Postsecretory modifications of streptavidin. *Biochem J* **259**, 369-376.
- Bayer, E.A., Ehrlich-Rogozinski, S., and Wilchek, M. (1996). Sodium dodecyl sulfate-polyacrylamide gel electrophoretic method for assessing the quaternary state and comparative thermostability of avidin and streptavidin. *Electrophoresis* **17**, 1319-1324.
- Beckett, D., Kovaleva, E., and Schatz, P.J. (1999). A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. *Protein Sci* **8**, 921-929.
- Begg, K.J., Dewar, S.J., and Donachie, W.D. (1995). A new *Escherichia coli* cell division gene, ftsK. *J Bacteriol* **177**, 6211-6222.
- Berg, O.G., and von Hippel, P.H. (1985). Diffusion-controlled macromolecular interactions. *Annu Rev Biophys Biophys Chem* **14**, 131-160.
- Bi, E.F., and Lutkenhaus, J. (1991). FtsZ ring structure associated with division in *Escherichia coli*. *Nature* **354**, 161-164.
- Bigot, S., Corre, J., Louarn, J.M., Cornet, F., and Barre, F.X. (2004). FtsK activities in Xer recombination, DNA mobilization and cell division involve overlapping and separate domains of the protein. *Mol Microbiol* **54**, 876-886.
- Bigot, S., and Mariani, K.J. (2010). DNA chirality-dependent stimulation of topoisomerase IV activity by the C-terminal AAA+ domain of FtsK. *Nucleic Acids Res* **38**, 3031-3040.

- Bigot, S., Saleh, O.A., Cornet, F., Allemand, J.F., and Barre, F.X. (2006). Oriented loading of FtsK on KOPS. *Nat Struct Mol Biol* 13, 1026-1028.
- Bigot, S., Sivanathan, V., Possoz, C., Barre, F.X., and Cornet, F. (2007). FtsK, a literate chromosome segregation machine. *Mol Microbiol* 64, 1434-1441.
- Blois, T.M., Hong, H., Kim, T.H., and Bowie, J.U. (2009). Protein unfolding with a steric trap. *J Am Chem Soc* 131, 13914-13915.
- Bloom, J.D., Labthavikul, S.T., Otey, C.R., and Arnold, F.H. (2006). Protein stability promotes evolvability. *Proc Natl Acad Sci U S A* 103, 5869-5874.
- Boas, M.A. (1927). The Effect of Desiccation upon the Nutritive Properties of Egg-white. *Biochem J* 21, 712-724 711.
- Bock, V.D., Perciaccante, R., Jansen, T.P., Hiemstra, H., and van Maarseveen, J.H. (2006). Click chemistry as a route to cyclic tetrapeptide analogues: synthesis of cyclo-[Pro-Val-psi(triazole)-Pro-Tyr]. *Org Lett* 8, 919-922.
- Boder, E.T., Midelfort, K.S., and Wittrup, K.D. (2000). Directed evolution of antibody fragments with monovalent femtomolar antigen-binding affinity. *Proc Natl Acad Sci U S A* 97, 10701-10705.
- Boerman, O.C., van Schaijk, F.G., Oyen, W.J., and Corstens, F.H. (2003). Pretargeted radioimmunotherapy of cancer: progress step by step. *J Nucl Med* 44, 400-411.
- Branton, D., Deamer, D.W., Marziali, A., Bayley, H., Benner, S.A., Butler, T., Di Ventra, M., Garaj, S., Hibbs, A., Huang, X., *et al.* (2008). The potential and challenges of nanopore sequencing. *Nat Biotechnol* 26, 1146-1153.
- Braslavsky, I., Hebert, B., Kartalov, E., and Quake, S.R. (2003). Sequence information can be obtained from single DNA molecules. *Proc Natl Acad Sci U S A* 100, 3960-3964.
- Breitz, H.B., Weiden, P.L., Beaumier, P.L., Axworthy, D.B., Seiler, C., Su, F.M., Graves, S., Bryan, K., and Reno, J.M. (2000). Clinical optimization of pretargeted radioimmunotherapy with antibody-streptavidin conjugate and 90Y-DOTA-biotin. *J Nucl Med* 41, 131-140.
- Brizzard, B.L., Chubet, R.G., and Vizard, D.L. (1994). Immunoaffinity purification of FLAG epitope-tagged bacterial alkaline phosphatase using a novel monoclonal antibody and peptide elution. *Biotechniques* 16, 730-735.
- Bruneau, E., Sutter, D., Hume, R.I., and Akaaboune, M. (2005). Identification of nicotinic acetylcholine receptor recycling and its role in maintaining receptor density at the neuromuscular junction in vivo. *J Neurosci* 25, 9949-9959.
- Brunger, A.T. (1991). Simulated annealing in crystallography. *Annual Review of Physical Chemistry* 42, 197-223.

- Brunger, A.T. (1992). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472-475.
- Buranda, T., Lopez, G.P., Keij, J., Harris, R., and Sklar, L.A. (1999). Peptides, antibodies, and FRET on beads in flow cytometry: A model system using fluoresceinated and biotinylated beta-endorphin. *Cytometry* **37**, 21-31.
- Butlin, N.G., and Meares, C.F. (2006). Antibodies with infinite affinity: origins and applications. *Acc Chem Res* **39**, 780-787.
- Byrd, A.K., and Raney, K.D. (2004). Protein displacement by an assembly of helicase molecules aligned along single-stranded DNA. *Nat Struct Mol Biol* **11**, 531-538.
- Byun, E., Kim, J., Kang, S.M., Lee, H., and Bang, D. (2011). Surface PEGylation via native chemical ligation. *Bioconjug Chem* **22**, 4-8.
- Casey, J.R., Grinstein, S., and Orlowski, J. (2010). Sensors and regulators of intracellular pH. *Nat Rev Mol Cell Biol* **11**, 50-61.
- Caswell, K.K., Wilson, J.N., Bunz, U.H., and Murphy, C.J. (2003). Preferential end-to-end assembly of gold nanorods by biotin-streptavidin connectors. *J Am Chem Soc* **125**, 13914-13915.
- Cedrone, F., Menez, A., and Quemeneur, E. (2000). Tailoring new enzyme functions by rational redesign. *Curr Opin Struct Biol* **10**, 405-410.
- Chalet, L., Miller, T.W., Tausig, F., and Wolf, F.J. (1963). Antibiotic Msd-235. li. Separation and Purification of Synergistic Components. *Antimicrob Agents Chemother (Bethesda)* **161**, 28-32.
- Chalet, L., and Wolf, F.J. (1964). The Properties of Streptavidin, a Biotin-Binding Protein Produced by Streptomyces. *Arch Biochem Biophys* **106**, 1-5.
- Chan, D.C., and Anderson, A.C. (2006). Towards species-specific antifolates. *Curr Med Chem* **13**, 377-398.
- Chen, I., Howarth, M., Lin, W., and Ting, A.Y. (2005). Site-specific labeling of cell surface proteins with biophysical probes using biotin ligase. *Nat Methods* **2**, 99-104.
- Chen, I., and Ting, A.Y. (2005). Site-specific labeling of proteins with small molecules in live cells. *Curr Opin Biotechnol* **16**, 35-40.
- Chen, J.C., and Beckwith, J. (2001). FtsQ, FtsL and FtsI require FtsK, but not FtsN, for co-localization with FtsZ during Escherichia coli cell division. *Mol Microbiol* **42**, 395-413.

- Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12-21.
- Chilkoti, A., Schwartz, B.L., Smith, R.D., Long, C.J., and Stayton, P.S. (1995a). Engineered chimeric streptavidin tetramers as novel tools for bioseparations and drug delivery. *Biotechnology (N Y)* **13**, 1198-1204.
- Chilkoti, A., Stayton, P.S. (1995). Molecular Origins of the Slow Streptavidin-Biotin Dissociation Kinetics. *Journal of the American Chemical Society* **117**, 10622-10628.
- Chilkoti, A., Tan, P.H., and Stayton, P.S. (1995b). Site-directed mutagenesis studies of the high-affinity streptavidin-biotin complex: contributions of tryptophan residues 79, 108, and 120. *Proc Natl Acad Sci U S A* **92**, 1754-1758.
- Chivers, C.E., Crozat, E., Chu, C., Moy, V.T., Sherratt, D.J., and Howarth, M. (2010). A streptavidin variant with slower biotin dissociation and increased mechanostability. *Nat Methods* **7**, 391-393.
- Chivers, C.E., Koner, A.L., Lowe, E.D., and Howarth, M. (2011). How the biotin-streptavidin interaction was made even stronger: investigation via crystallography and a chimaeric tetramer. *Biochem J* **435**, 55-63.
- Chmura, A.J., Orton, M.S., and Meares, C.F. (2001). Antibodies with infinite affinity. *Proc Natl Acad Sci U S A* **98**, 8480-8484.
- Chu, V., Freitag, S., Le Trong, I., Stenkamp, R.E., and Stayton, P.S. (1998). Thermodynamic and structural consequences of flexible loop deletion by circular permutation in the streptavidin-biotin system. *Protein Sci* **7**, 848-859.
- Cleland, J.L., Builder, S.E., Swartz, J.R., Winkler, M., Chang, J.Y., and Wang, D.I. (1992). Polyethylene glycol enhanced protein refolding. *Biotechnology (N Y)* **10**, 1013-1019.
- Corazza, A., Rosano, C., Pagano, K., Alverdi, V., Esposito, G., Capanni, C., Bemporad, F., Plakoutsi, G., Stefani, M., Chiti, F., *et al.* (2006). Structure, conformational stability, and enzymatic properties of acylphosphatase from the hyperthermophile *Sulfolobus solfataricus*. *Proteins* **62**, 64-79.
- Crozat, E., and Grainge, I. (2010). FtsK DNA translocase: the fast motor that knows where it's going. *Chembiochem* **11**, 2232-2243.
- Crozat, E., Meglio, A., Allemand, J.F., Chivers, C.E., Howarth, M., Venien-Bryan, C., Grainge, I., and Sherratt, D.J. (2010). Separating speed and ability to displace roadblocks during DNA translocation by FtsK. *EMBO J* **29**, 1423-1433.
- Das, R., and Baker, D. (2008). Macromolecular modeling with rosetta. *Annu Rev Biochem* **77**, 363-382.

- Davis, G.D., Elisee, C., Newham, D.M., and Harrison, R.G. (1999). New fusion protein systems designed to give soluble expression in *Escherichia coli*. *Biotechnol Bioeng* 65, 382-388.
- Dawson, P.E., Muir, T.W., Clark-Lewis, I., and Kent, S.B. (1994). Synthesis of proteins by native chemical ligation. *Science* 266, 776-779.
- Day, R.N., and Davidson, M.W. (2009). The fluorescent protein palette: tools for cellular imaging. *Chem Soc Rev* 38, 2887-2921.
- de Boer, E., Rodriguez, P., Bonte, E., Krijgsveld, J., Katsantoni, E., Heck, A., Grosveld, F., and Strouboulis, J. (2003). Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice. *Proc Natl Acad Sci U S A* 100, 7480-7485.
- DeChancie, J., and Houk, K.N. (2007). The origins of femtomolar protein-ligand binding: hydrogen-bond cooperativity and desolvation energetics in the biotin-(strept)avidin binding site. *J Am Chem Soc* 129, 5419-5429.
- Di Lallo, G., Fagioli, M., Barionovi, D., Ghelardini, P., and Paolozzi, L. (2003). Use of a two-hybrid assay to study the assembly of a complex multicomponent protein machinery: bacterial septosome differentiation. *Microbiology* 149, 3353-3359.
- Dietrich, J.A., McKee, A.E., and Keasling, J.D. (2010). High-throughput metabolic engineering: advances in small-molecule screening and selection. *Annu Rev Biochem* 79, 563-590.
- Ding, Z., Fong, R.B., Long, C.J., Stayton, P.S., and Hoffman, A.S. (2001). Size-dependent control of the binding of biotinylated proteins to streptavidin using a polymer shield. *Nature* 411, 59-62.
- Dixon, R.W., Radmer, R.J., Kuhn, B., Kollman, P.A., Yang, J., Raposo, C., Wilcox, C.S., Klumb, L.A., Stayton, P.S., Behnke, C., *et al.* (2002). Theoretical and experimental studies of biotin analogues that bind almost as tightly to streptavidin as biotin. *J Org Chem* 67, 1827-1837.
- Donovan, J.W., and Ross, K.D. (1973). Increase in the stability of avidin produced by binding of biotin. A differential scanning calorimetric study of denaturation by heat. *Biochemistry* 12, 512-517.
- Draper, G.C., McLennan, N., Begg, K., Masters, M., and Donachie, W.D. (1998). Only the N-terminal domain of FtsK functions in cell division. *J Bacteriol* 180, 4621-4627.
- Dressman, D., Yan, H., Traverso, G., Kinzler, K.W., and Vogelstein, B. (2003). Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. *Proc Natl Acad Sci U S A* 100, 8817-8822.
- du Vigneaud, V. (1942). The Structure of Biotin. *Science* 96, 455-461.

- Eakin, R.E., McKinley, W.A., and Williams, R.J. (1940a). Egg-White Injury in Chicks and Its Relationship to a Deficiency of Vitamin H (Biotin). *Science* **92**, 224-225.
- Eakin, R.E., Snell, E.E., and Williams, R.J. (1940b). A constituent of raw egg white capable of inactivating biotin in vitro. *J Biol Chem* **140**, 535-543.
- Eakin, R.E., Snell, E.E., and Williams, R.J. (1941). The concentration and assay of avidin, the injury-producing protein in raw egg white. *J Biol Chem* **140**, 535-543.
- Eisenberg-Domovich, Y., Hytonen, V.P., Wilchek, M., Bayer, E.A., Kulomaa, M.S., and Livnah, O. (2005). High-resolution crystal structure of an avidin-related protein: insight into high-affinity biotin binding and protein stability. *Acta Crystallogr D Biol Crystallogr* **61**, 528-538.
- el Hawrani, A.S., Sessions, R.B., Moreton, K.M., and Holbrook, J.J. (1996). Guided evolution of enzymes with new substrate specificities. *J Mol Biol* **264**, 97-110.
- Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132.
- Esvelt, K.M., Carlson, J.C., and Liu, D.R. (2011). A system for the continuous directed evolution of biomolecules. *Nature* **472**, 499-503.
- Evans, E., and Ritchie, K. (1997). Dynamic strength of molecular adhesion bonds. *Biophys J* **72**, 1541-1555.
- Fall, R.R., Nervi, A.M., Alberts, A.W., and Vagelos, P.R. (1971). Acetyl CoA carboxylase: isolation and characterization of native biotin carboxyl carrier protein. *Proc Natl Acad Sci U S A* **68**, 1512-1515.
- Farre, A., van der Horst, A., Blab, G.A., Downing, B.P., and Forde, N.R. (2010). Stretching single DNA molecules to demonstrate high-force capabilities of holographic optical tweezers. *J Biophotonics* **3**, 224-233.
- Florin, E.L., Moy, V.T., and Gaub, H.E. (1994). Adhesion forces between individual ligand-receptor pairs. *Science* **264**, 415-417.
- Foote, J., and Eisen, H.N. (2000). Breaking the affinity ceiling for antibodies and T cell receptors. *Proc Natl Acad Sci U S A* **97**, 10679-10681.
- Francis, D.M., and Page, R. (2010). Strategies to optimize protein expression in *E. coli*. *Curr Protoc Protein Sci* *Chapter 5*, Unit 5 24 21-29.
- Freitag, S., Chu, V., Penzotti, J.E., Klumb, L.A., To, R., Hyre, D., Le Trong, I., Lybrand, T.P., Stenkamp, R.E., and Stayton, P.S. (1999). A structural snapshot of an intermediate on the streptavidin-biotin dissociation pathway. *Proc Natl Acad Sci U S A* **96**, 8384-8389.

- Freitag, S., Le Trong, I., Klumb, L., Stayton, P.S., and Stenkamp, R.E. (1997). Structural studies of the streptavidin binding loop. *Protein Sci* 6, 1157-1166.
- French, S., and Wilson, K. (1978). On the treatment of negative intensity observations. *Acta Crystallographica Section A* 34, 517-525.
- Fudem-Goldin, B., and Orr, G.A. (1990). 2-Iminobiotin-containing reagent and affinity columns. *Methods Enzymol* 184, 167-173.
- Gallizia, A., de Lalla, C., Nardone, E., Santambrogio, P., Brandazza, A., Sidoli, A., and Arosio, P. (1998). Production of a soluble and functional recombinant streptavidin in *Escherichia coli*. *Protein Expr Purif* 14, 192-196.
- Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., and Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. In *The Proteomics Protocols Handbook*, J.M. Walker, ed. (Humana Press Inc.), pp. 571-607.
- Giebel, L.B., Cass, R.T., Milligan, D.L., Young, D.C., Arze, R., and Johnson, C.R. (1995). Screening of cyclic peptide phage libraries identifies ligands that bind streptavidin with high affinities. *Biochemistry* 34, 15430-15435.
- Gonzalez, M., Argarana, C.E., and Fidelio, G.D. (1999). Extremely high thermal stability of streptavidin and avidin upon biotin binding. *Biomol Eng* 16, 67-72.
- Graham, J.E., Sivanathan, V., Sherratt, D.J., and Arciszewska, L.K. (2010). FtsK translocation on DNA stops at XerCD-dif. *Nucleic Acids Res* 38, 72-81.
- Grana, C., Chinol, M., Robertson, C., Mazzetta, C., Bartolomei, M., De Cicco, C., Fiorenza, M., Gatti, M., Caliceti, P., and Paganelli, G. (2002). Pretargeted adjuvant radioimmunotherapy with yttrium-90-biotin in malignant glioma patients: a pilot study. *Br J Cancer* 86, 207-212.
- Green, N.M. (1963). Avidin. 1. The Use of (14-C)Biotin for Kinetic Studies and for Assay. *Biochem J* 89, 585-591.
- Green, N.M. (1966). Thermodynamics of the binding of biotin and some analogues by avidin. *Biochem J* 101, 774-780.
- Green, N.M. (1970). Spectrophotometric determination of avidin and biotin. *Methods Enzymol* 18, 418-424.
- Green, N.M. (1975). Avidin. *Adv Protein Chem* 29, 85-133.
- Green, N.M. (1990). Avidin and streptavidin. *Methods Enzymol* 184, 51-67.
- Green, N.M., and Toms, E.J. (1973). The properties of subunits of avidin coupled to sepharose. *Biochem J* 133, 687-700.

- Gribenko, A.V., Patel, M.M., Liu, J., McCallum, S.A., Wang, C., and Makhatadze, G.I. (2009). Rational stabilization of enzymes by computational redesign of surface charge-charge interactions. *Proc Natl Acad Sci U S A* **106**, 2601-2606.
- Griffin, B.A., Adams, S.R., Jones, J., and Tsien, R.Y. (2000). Fluorescent labeling of recombinant proteins in living cells with FIAsH. *Methods Enzymol* **327**, 565-578.
- Griffin, B.A., Adams, S.R., and Tsien, R.Y. (1998). Specific covalent labeling of recombinant protein molecules inside live cells. *Science* **281**, 269-272.
- Grubmuller, H., Heymann, B., and Tavan, P. (1996). Ligand binding: molecular mechanics calculation of the streptavidin-biotin rupture force. *Science* **271**, 997-999.
- Gyorgy, P. (1939). The curative factor (vitamin H) for egg white injury, with particular reference to its presence in different foodstuffs and in yeast. *Journal of Biological Chemistry* **131**, 733-744.
- Gyorgy, P., Rose, C.S., Hofmann, K., Melville, D.B., and V, D.U.V. (1940). A Further Note on the Identity of Vitamin H with Biotin. *Science* **92**, 609.
- Harry, E., Monahan, L., and Thompson, L. (2006). Bacterial cell division: the mechanism and its precision. *Int Rev Cytol* **253**, 27-94.
- Hayouka, R., Eisenberg-Domovich, Y., Hytonen, V.P., Maatta, J.A., Nordlund, H.R., Kulomaa, M.S., Wilchek, M., Bayer, E.A., and Livnah, O. (2008). Critical importance of loop conformation to avidin-enhanced hydrolysis of an active biotin ester. *Acta Crystallogr D Biol Crystallogr* **64**, 302-308.
- Healy, S., Perez-Cadahia, B., Jia, D., McDonald, M.K., Davie, J.R., and Gravel, R.A. (2009). Biotin is not a natural histone modification. *Biochim Biophys Acta* **1789**, 719-733.
- Hein, C.D., Liu, X.M., and Wang, D. (2008). Click chemistry, a powerful tool for pharmaceutical sciences. *Pharm Res* **25**, 2216-2230.
- Helppolainen, S.H., Maatta, J.A., Halling, K.K., Slotte, J.P., Hytonen, V.P., Janis, J., Vainiotalo, P., Kulomaa, M.S., and Nordlund, H.R. (2008). Bradavidin II from *Bradyrhizobium japonicum*: a new avidin-like biotin-binding protein. *Biochim Biophys Acta* **1784**, 1002-1010.
- Helppolainen, S.H., Nurminen, K.P., Maatta, J.A., Halling, K.K., Slotte, J.P., Huhtala, T., Liimatainen, T., Yla-Herttuala, S., Airene, K.J., Narvanen, A., *et al.* (2007). Rhizavidin from *Rhizobium etli*: the first natural dimer in the avidin protein family. *Biochem J* **405**, 397-405.
- Hendrickson, W.A., Pahler, A., Smith, J.L., Satow, Y., Merritt, E.A., and Phizackerley, R.P. (1989). Crystal structure of core streptavidin determined from multiwavelength anomalous diffraction of synchrotron radiation. *Proc Natl Acad Sci U S A* **86**, 2190-2194.

- Hiller, Y., Bayer, E.A., and Wilchek, M. (1990). Nonglycosylated avidin. *Methods Enzymol* **184**, 68-70.
- Hirsch, J.D., Eslamizar, L., Filanoski, B.J., Malekzadeh, N., Haugland, R.P., and Beechem, J.M. (2002). Easily reversible desthiobiotin binding to streptavidin, avidin, and other biotin-binding proteins: uses for protein labeling, detection, and isolation. *Anal Biochem* **308**, 343-357.
- Hofmann, K., Titus, G., Montibeller, J.A., and Finn, F.M. (1982). Avidin binding of carboxyl-substituted biotin and analogues. *Biochemistry* **21**, 978-984.
- Hofmann, K., Wood, S.W., Brinton, C.C., Montibeller, J.A., and Finn, F.M. (1980). Iminobiotin affinity columns and their application to retrieval of streptavidin. *Proc Natl Acad Sci U S A* **77**, 4666-4668.
- Hohsaka, T., Muranaka, N., Komiyama, C., Matsui, K., Takaura, S., Abe, R., Murakami, H., and Sisido, M. (2004). Position-specific incorporation of dansylated non-natural amino acids into streptavidin by using a four-base codon. *FEBS Lett* **560**, 173-177.
- Holm, L., Moody, P., and Howarth, M. (2009). Electrophilic affibodies forming covalent bonds to protein targets. *J Biol Chem* **284**, 32906-32913.
- Holmberg, A., Blomstergren, A., Nord, O., Lukacs, M., Lundeberg, J., and Uhlen, M. (2005). The biotin-streptavidin interaction can be reversibly broken using water at elevated temperatures. *Electrophoresis* **26**, 501-510.
- Hong, H., Blois, T.M., Cao, Z., and Bowie, J.U. (2010). Method to measure strong protein-protein interactions in lipid bilayers using a steric trap. *Proc Natl Acad Sci U S A* **107**, 19802-19807.
- Howarth, M., Chinnapen, D.J., Gerrow, K., Dorrestein, P.C., Grandy, M.R., Kelleher, N.L., El-Husseini, A., and Ting, A.Y. (2006). A monovalent streptavidin with a single femtomolar biotin binding site. *Nat Methods* **3**, 267-273.
- Howarth, M., Liu, W., Puthenveetil, S., Zheng, Y., Marshall, L.F., Schmidt, M.M., Wittrup, K.D., Bawendi, M.G., and Ting, A.Y. (2008). Monovalent, reduced-size quantum dots for imaging receptors on living cells. *Nat Methods* **5**, 397-399.
- Howarth, M., Takao, K., Hayashi, Y., and Ting, A.Y. (2005). Targeting quantum dots to surface proteins in living cells with biotin ligase. *Proc Natl Acad Sci U S A* **102**, 7583-7588.
- Howarth, M., and Ting, A.Y. (2008). Imaging proteins in live mammalian cells with biotin ligase and monovalent streptavidin. *Nat Protoc* **3**, 534-545.
- Huang, J., Zarnitsyna, V.I., Liu, B., Edwards, L.J., Jiang, N., Evavold, B.D., and Zhu, C. (2010). The kinetics of two-dimensional TCR and pMHC interactions determine T-cell responsiveness. *Nature* **464**, 932-936.

- Huberman, T., Eisenberg-Domovich, Y., Gitlin, G., Kulik, T., Bayer, E.A., Wilchek, M., and Livnah, O. (2001). Chicken avidin exhibits pseudo-catalytic properties. Biochemical, structural, and electrostatic consequences. *J Biol Chem* 276, 32031-32039.
- Hudson, P.J., and Souriau, C. (2003). Engineered antibodies. *Nat Med* 9, 129-134.
- Huisgen, R. (1963). Kinetics and Mechanism of 1,3-Dipolar Cycloadditions. *Angew Chem Int Ed Engl* 2, 633-696.
- Humbert, N., Schurmann, P., Zocchi, A., Neuhaus, J.M., and Ward, T.R. (2008). High-yield production and purification of recombinant T7-tag mature streptavidin in glucose-stressed *E. coli*. *Methods Mol Biol* 418, 101-110.
- Hurley, W.L., Finkelstein, E., and Holst, B.D. (1985). Identification of surface proteins on bovine leukocytes by a biotin-avidin protein blotting technique. *J Immunol Methods* 85, 195-202.
- Hutter, J.L., and Bechhoefer, J. (1993). Calibration of Atomic-Force Microscope Tips. *Rev Sci Instrum* 64, 1868-1873.
- Hymes, J., and Wolf, B. (1996). Biotinidase and its roles in biotin metabolism. *Clin Chim Acta* 255, 1-11.
- Hyre, D.E., Amon, L.M., Penzotti, J.E., Le Trong, I., Stenkamp, R.E., Lybrand, T.P., and Stayton, P.S. (2002). Early mechanistic events in biotin dissociation from streptavidin. *Nat Struct Biol* 9, 582-585.
- Hyre, D.E., Le Trong, I., Freitag, S., Stenkamp, R.E., and Stayton, P.S. (2000). Ser45 plays an important role in managing both the equilibrium and transition state energetics of the streptavidin-biotin system. *Protein Sci* 9, 878-885.
- Hyre, D.E., Le Trong, I., Merritt, E.A., Eccleston, J.F., Green, N.M., Stenkamp, R.E., and Stayton, P.S. (2006). Cooperative hydrogen bond interactions in the streptavidin-biotin system. *Protein Sci* 15, 459-467.
- Hytonen, V.P., Laitinen, O.H., Grapputo, A., Kettunen, A., Savolainen, J., Kalkkinen, N., Marttila, A.T., Nordlund, H.R., Nyholm, T.K., Paganelli, G., *et al.* (2003). Characterization of poultry egg-white avidins and their potential as a tool in pretargeting cancer treatment. *Biochem J* 372, 219-225.
- Hytonen, V.P., Maatta, J.A., Kidron, H., Halling, K.K., Horha, J., Kulomaa, T., Nyholm, T.K., Johnson, M.S., Salminen, T.A., Kulomaa, M.S., *et al.* (2005). Avidin related protein 2 shows unique structural and functional features among the avidin protein family. *BMC Biotechnol* 5, 28.

- Hytonen, V.P., Nyholm, T.K., Pentikainen, O.T., Vaarno, J., Porkka, E.J., Nordlund, H.R., Johnson, M.S., Slotte, J.P., Laitinen, O.H., and Kulomaa, M.S. (2004). Chicken avidin-related protein 4/5 shows superior thermal stability when compared with avidin while retaining high affinity to biotin. *J Biol Chem* 279, 9337-9343.
- Jackel, C., Kast, P., and Hilvert, D. (2008). Protein design by directed evolution. *Annu Rev Biophys* 37, 153-173.
- James, L.C., and Tawfik, D.S. (2003). The specificity of cross-reactivity: promiscuous antibody binding involves specific hydrogen bonds rather than nonspecific hydrophobic stickiness. *Protein Sci* 12, 2183-2193.
- Janolino, V.G., Fontecha, J., and Swaisgood, H.E. (1996). A spectrophotometric assay for biotin binding sites of immobilized avidin. *Applied Biochemistry and Biotechnology* 56, 1-7.
- Jen, M., and Yan, A.C. (2010). Syndromes associated with nutritional deficiency and excess. *Clin Dermatol* 28, 669-685.
- Jones, M.L., and Kurzban, G.P. (1995). Noncooperativity of biotin binding to tetrameric streptavidin. *Biochemistry* 34, 11750-11756.
- Kada, G., Falk, H., and Gruber, H.J. (1999). Accurate measurement of avidin and streptavidin in crude biofluids with a new, optimized biotin-fluorescein conjugate. *Biochim Biophys Acta* 1427, 33-43.
- Kapust, R.B., and Waugh, D.S. (1999). Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci* 8, 1668-1674.
- Katz, B.A. (1995). Binding to protein targets of peptidic leads discovered by phage display: crystal structures of streptavidin-bound linear and cyclic peptide ligands containing the HPQ sequence. *Biochemistry* 34, 15421-15429.
- Katz, B.A. (1997). Binding of biotin to streptavidin stabilizes intersubunit salt bridges between Asp61 and His87 at low pH. *J Mol Biol* 274, 776-800.
- Kesapragada, S.V., Yim, T.J., Dordick, J.S., Kane, R.S., and Gall, D. (2010). Selective assembly of multi-component nanosprings and nanorods. *J Nanosci Nanotechnol* 10, 2252-2256.
- Klumb, L.A., Chu, V., and Stayton, P.S. (1998). Energetic roles of hydrogen bonds at the ureido oxygen binding pocket in the streptavidin-biotin complex. *Biochemistry* 37, 7657-7663.
- Kobayashi, N., and Oyama, H. (2011). Antibody engineering toward high-sensitivity high-throughput immunosensing of small molecules. *Analyst* 136, 642-651.

- Kolb, H.C., Finn, M.G., and Sharpless, K.B. (2001). Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew Chem Int Ed Engl* **40**, 2004-2021.
- Kolomiets, E.I., and Zdor, N.A. (1998). Streptavidin, a product of the strain *Streptomyces avidinii* VKM Ac1047: synthesis, purification and use in immunoassay technology. *Russian Journal of Biotechnology* **2**, 1-8.
- Kranz, D.M., Herron, J.N., and Voss, E.W., Jr. (1982). Mechanisms of ligand binding by monoclonal anti-fluorescein antibodies. *J Biol Chem* **257**, 6987-6995.
- Kuhn, B., and Kollman, P.A. (2000). A ligand that is predicted to bind better to avidin than biotin: insights from computational fluorine scanning. *J Am Chem Soc* **122**, 3909 - 3916.
- Kuntz, I.D., Chen, K., Sharp, K.A., and Kollman, P.A. (1999). The maximal affinity of ligands. *Proc Natl Acad Sci U S A* **96**, 9997-10002.
- Kyratsous, C.A., Silverstein, S.J., DeLong, C.R., and Panagiotidis, C.A. (2009). Chaperone-fusion expression plasmid vectors for improved solubility of recombinant proteins in *Escherichia coli*. *Gene* **440**, 9-15.
- Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685.
- Laitinen, O.H., Airenne, K.J., Marttila, A.T., Kulik, T., Porkka, E., Bayer, E.A., Wilchek, M., and Kulomaa, M.S. (1999). Mutation of a critical tryptophan to lysine in avidin or streptavidin may explain why sea urchin fibropellin adopts an avidin-like domain. *FEBS Lett* **461**, 52-58.
- Laitinen, O.H., Hytonen, V.P., Ahlroth, M.K., Pentikainen, O.T., Gallagher, C., Nordlund, H.R., Ovod, V., Marttila, A.T., Porkka, E., Heino, S., *et al.* (2002). Chicken avidin-related proteins show altered biotin-binding and physico-chemical properties as compared with avidin. *Biochem J* **363**, 609-617.
- Laitinen, O.H., Hytonen, V.P., Nordlund, H.R., and Kulomaa, M.S. (2006). Genetically engineered avidins and streptavidins. *Cell Mol Life Sci* **63**, 2992-3017.
- Laitinen, O.H., Nordlund, H.R., Hytonen, V.P., and Kulomaa, M.S. (2007). Brave new (strept)avidins in biotechnology. *Trends Biotechnol* **25**, 269-277.
- Lamla, T., and Erdmann, V.A. (2004). The Nano-tag, a streptavidin-binding peptide for the purification and detection of recombinant proteins. *Protein Expr Purif* **33**, 39-47.
- Lau, I.F., Filipe, S.R., Soballe, B., Okstad, O.A., Barre, F.X., and Sherratt, D.J. (2003). Spatial and temporal organization of replicating *Escherichia coli* chromosomes. *Mol Microbiol* **49**, 731-743.

- Laursen, B.S., Mortensen, K.K., Sperling-Petersen, H.U., and Hoffman, D.W. (2003). A conserved structural motif at the N terminus of bacterial translation initiation factor IF2. *J Biol Chem* 278, 16320-16328.
- LaVallie, E.R., Lu, Z., Diblasio-Smith, E.A., Collins-Racie, L.A., and McCoy, J.M. (2000). Thioredoxin as a fusion partner for production of soluble recombinant proteins in *Escherichia coli*. *Methods Enzymol* 326, 322-340.
- Le Trong, I., Humbert, N., Ward, T.R., and Stenkamp, R.E. (2006). Crystallographic analysis of a full-length streptavidin with its C-terminal polypeptide bound in the biotin binding site. *J Mol Biol* 356, 738-745.
- Lease, J.G., Parsons, H.T., and Kelly, E. (1937). A comparison in five types of animals of the effects of dietary egg white and of a specific factor given orally or parenterally. *Biochem J* 31, 433-437.
- Leon-Del-Rio, A., Leclerc, D., Akerman, B., Wakamatsu, N., and Gravel, R.A. (1995). Isolation of a cDNA encoding human holocarboxylase synthetase by functional complementation of a biotin auxotroph of *Escherichia coli*. *Proc Natl Acad Sci U S A* 92, 4626-4630.
- Levy, M., and Ellington, A.D. (2008). Directed evolution of streptavidin variants using in vitro compartmentalization. *Chem Biol* 15, 979-989.
- Li, Q., Gusarov, S., Evoy, S., and Kovalenko, A. (2009). Electronic structure, binding energy, and solvation structure of the streptavidin-biotin supramolecular complex: ONIOM and 3D-RISM study. *J Phys Chem B* 113, 9958-9967.
- Lin, C.W., and Ting, A.Y. (2006). Transglutaminase-catalyzed site-specific conjugation of small-molecule probes to proteins in vitro and on the surface of living cells. *J Am Chem Soc* 128, 4542-4543.
- Link, A.J., and Tirrell, D.A. (2003). Cell surface labeling of *Escherichia coli* via copper(I)-catalyzed [3+2] cycloaddition. *J Am Chem Soc* 125, 11164-11165.
- Livnah, O., Bayer, E.A., Wilchek, M., and Sussman, J.L. (1993). Three-dimensional structures of avidin and the avidin-biotin complex. *Proc Natl Acad Sci U S A* 90, 5076-5080.
- Lopez-Calderero, I., Sanchez Chavez, E., and Garcia-Carbonero, R. (2010). The insulin-like growth factor pathway as a target for cancer therapy. *Clin Transl Oncol* 12, 326-338.
- Lorimer, G.H. (1996). A quantitative assessment of the role of the chaperonin proteins in protein folding in vivo. *FASEB J* 10, 5-9.
- Los, G.V., Encell, L.P., McDougall, M.G., Hartzell, D.D., Karassina, N., Zimprich, C., Wood, M.G., Learish, R., Ohana, R.F., Urh, M., *et al.* (2008). HaloTag: a novel protein labeling technology for cell imaging and protein analysis. *ACS Chem Biol* 3, 373-382.

- Lowe, J., Ellonen, A., Allen, M.D., Atkinson, C., Sherratt, D.J., and Grainge, I. (2008). Molecular mechanism of sequence-directed DNA loading and translocation by FtsK. *Mol Cell* 31, 498-509.
- Maatta, J.A., Helppolainen, S.H., Hytonen, V.P., Johnson, M.S., Kulomaa, M.S., Airenne, T.T., and Nordlund, H.R. (2009). Structural and functional characteristics of xenavidin, the first frog avidin from *Xenopus tropicalis*. *BMC Struct Biol* 9, 63.
- Mach-Aigner, A.R., Grosstessner-Hain, K., Pocas-Fonseca, M.J., Mechtler, K., and Mach, R.L. (2010). From an electrophoretic mobility shift assay to isolated transcription factors: a fast genomic-proteomic approach. *BMC Genomics* 11, 644.
- Magalhaes, M.L., Czekster, C.M., Guan, R., Malashkevich, V.N., Almo, S.C., and Levy, M. (2011). Evolved streptavidin mutants reveal key role of loop residue in high-affinity binding. *Protein Sci* 20, 1145-1154.
- Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., Chen, Z., *et al.* (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437, 376-380.
- Markwick, N.P., Docherty, L.C., Phung, M.M., Lester, M.T., Murray, C., Yao, J.L., Mitra, D.S., Cohen, D., Beuning, L.L., Kutty-Amma, S., *et al.* (2003). Transgenic tobacco and apple plants expressing biotin-binding proteins are resistant to two cosmopolitan insect pests, potato tuber moth and lightbrown apple moth, respectively. *Transgenic Res* 12, 671-681.
- Marshall, D., Pedley, R.B., Boden, J.A., Boden, R., Melton, R.G., and Begent, R.H. (1996). Polyethylene glycol modification of a galactosylated streptavidin clearing agent: effects on immunogenicity and clearance of a biotinylated anti-tumour antibody. *Br J Cancer* 73, 565-572.
- Marvin, J.S., and Lowman, H.B. (2003). Redesigning an antibody fragment for faster association with its antigen. *Biochemistry* 42, 7077-7083.
- Massey, T.H., Mercogliano, C.P., Yates, J., Sherratt, D.J., and Lowe, J. (2006). Double-stranded DNA translocation: structure and mechanism of hexameric FtsK. *Mol Cell* 23, 457-469.
- Massoud, T.F., and Gambhir, S.S. (2003). Molecular imaging in living subjects: seeing fundamental biological processes in a new light. *Genes Dev* 17, 545-580.
- Matsumoto, T., Sawamoto, S., Sakamoto, T., Tanaka, T., Fukuda, H., and Kondo, A. (2011). Site-specific tetrameric streptavidin-protein conjugation using sortase A. *J Biotechnol* 152, 37-42.
- McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. *J Appl Crystallogr* 40, 658-674.

- McInnes, J.L., Forster, A.C., Skingle, D.C., and Symons, R.H. (1990). Preparation and uses of photobiotin. *Methods Enzymol* 184, 588-600.
- Meir, A., Helppolainen, S.H., Podoly, E., Nordlund, H.R., Hytonen, V.P., Maatta, J.A., Wilchek, M., Bayer, E.A., Kulomaa, M.S., and Livnah, O. (2009). Crystal structure of rhizavidin: insights into the enigmatic high-affinity interaction of an innate biotin-binding protein dimer. *J Mol Biol* 386, 379-390.
- Meng, X., Yang, X., Wang, K., Guo, Q., Tan, Y., Mo, Q., and Xu, X. (2010). Direct fluorescence detection of point mutations in human genomic DNA using microbead-based ligase chain reaction. *Talanta* 80, 1725-1729.
- Mergulhao, F.J., Summers, D.K., and Monteiro, G.A. (2005). Recombinant protein secretion in *Escherichia coli*. *Biotechnol Adv* 23, 177-202.
- Merkel, R., Nassoy, P., Leung, A., Ritchie, K., and Evans, E. (1999). Energy landscapes of receptor-ligand bonds explored with dynamic force spectroscopy. *Nature* 397, 50-53.
- Meyer, D.L., Schultz, J., Lin, Y., Henry, A., Sanderson, J., Jackson, J.M., Goshorn, S., Rees, A.R., and Graves, S.S. (2001). Reduced antibody response to streptavidin through site-directed mutagenesis. *Protein Sci* 10, 491-503.
- Midelfort, K.S., Hernandez, H.H., Lippow, S.M., Tidor, B., Drennan, C.L., and Wittrup, K.D. (2004). Substantial energetic improvement with minimal structural perturbation in a high affinity mutant antibody. *J Mol Biol* 343, 685-701.
- Miksch, G., Ryu, S., Risse, J.M., and Flaschel, E. (2008). Factors that influence the extracellular expression of streptavidin in *Escherichia coli* using a bacteriocin release protein. *Appl Microbiol Biotechnol* 81, 319-326.
- Miller, L.W., Sable, J., Goelet, P., Sheetz, M.P., and Cornish, V.W. (2004). Methotrexate conjugates: a molecular in vivo protein tag. *Angew Chem Int Ed Engl* 43, 1672-1675.
- Miyamoto, S., and Kollman, P.A. (1993). Absolute and relative binding free energy calculations of the interaction of biotin and its analogs with streptavidin using molecular dynamics/free energy perturbation approaches. *Proteins* 16, 226-245.
- Mock, D.M., and Heird, G.M. (1997). Urinary biotin analogs increase in humans during chronic supplementation: the analogs are biotin metabolites. *Am J Physiol* 272, E83-85.
- Morris, P.D., and Raney, K.D. (1999). DNA helicases displace streptavidin from biotin-labeled oligonucleotides. *Biochemistry* 38, 5164-5171.
- Morris, P.D., Tackett, A.J., and Raney, K.D. (2001). Biotin-streptavidin-labeled oligonucleotides as probes of helicase mechanisms. *Methods* 23, 149-159.

- Moss, J., and Lane, M.D. (1971). The biotin-dependent enzymes. *Adv Enzymol Relat Areas Mol Biol* 35, 321-442.
- Moy, V.T., Florin, E.L., and Gaub, H.E. (1994). Intermolecular forces and energies between ligands and receptors. *Science* 266, 257-259.
- Murray, S., Maraveyas, A., Dougan, T., and Chu, A.C. (2002). Reduction of streptavidin RYDS-mediated renal adhesion by site-directed mutagenesis. *Biochim Biophys Acta* 1570, 81-88.
- Nagarajan, V., Ramaley, R., Albertson, H., and Chen, M. (1993). Secretion of streptavidin from *Bacillus subtilis*. *Appl Environ Microbiol* 59, 3894-3898.
- Nordlund, H.R., Hytonen, V.P., Horha, J., Maatta, J.A., White, D.J., Halling, K., Porkka, E.J., Slotte, J.P., Laitinen, O.H., and Kulomaa, M.S. (2005a). Tetravalent single-chain avidin: from subunits to protein domains via circularly permuted avidins. *Biochem J* 392, 485-491.
- Nordlund, H.R., Hytonen, V.P., Laitinen, O.H., and Kulomaa, M.S. (2005b). Novel avidin-like protein from a root nodule symbiotic bacterium, *Bradyrhizobium japonicum*. *J Biol Chem* 280, 13250-13255.
- Nordlund, H.R., Laitinen, O.H., Uotila, S.T., Nyholm, T., Hytonen, V.P., Slotte, J.P., and Kulomaa, M.S. (2003). Enhancing the thermal stability of avidin. Introduction of disulfide bridges between subunit interfaces. *J Biol Chem* 278, 2479-2483.
- Noren, C.J., Anthony-Cahill, S.J., Griffith, M.C., and Schultz, P.G. (1989). A general method for site-specific incorporation of unnatural amino acids into proteins. *Science* 244, 182-188.
- Northrup, S.H., and Erickson, H.P. (1992). Kinetics of protein-protein association explained by Brownian dynamics computer simulation. *Proc Natl Acad Sci U S A* 89, 3338-3342.
- Offman, M.N., Krol, M., Patel, N., Krishnan, S., Liu, J., Saha, V., and Bates, P.A. (2011). Rational engineering of L-asparaginase reveals importance of dual activity for cancer cell toxicity. *Blood* 117, 1614-1621.
- Ohana, R.F., Hurst, R., Vidugiriene, J., Slater, M.R., Wood, K.V., and Urh, M. (2011). HaloTag-based purification of functional human kinases from mammalian cells. *Protein Expr Purif* 76, 154-164.
- Orr, G.A. (1981). The use of the 2-iminobiotin-avidin interaction for the selective retrieval of labeled plasma membrane components. *J Biol Chem* 256, 761-766.
- Pahler, A., Hendrickson, W.A., Kolks, M.A., Argarana, C.E., and Cantor, C.R. (1987). Characterization and crystallization of core streptavidin. *J Biol Chem* 262, 13933-13937.

- Payne, G.F., DelaCruz, N., and Coppella, S.J. (1990). Improved production of heterologous protein from *Streptomyces lividans*. *Appl Microbiol Biotechnol* 33, 395-400.
- Pazy, Y., Kulik, T., Bayer, E.A., Wilchek, M., and Livnah, O. (2002). Ligand exchange between proteins. Exchange of biotin and biotin derivatives between avidin and streptavidin. *J Biol Chem* 277, 30892-30900.
- Pease, P.J., Levy, O., Cost, G.J., Gore, J., Ptacin, J.L., Sherratt, D., Bustamante, C., and Cozzarelli, N.R. (2005). Sequence-directed DNA translocation by purified FtsK. *Science* 307, 586-590.
- Petronzelli, F., Pelliccia, A., Anastasi, A.M., Lindstedt, R., Manganello, S., Ferrari, L.E., Albertoni, C., Leoni, B., Rosi, A., D'Alessio, V., *et al.* (2010). Therapeutic use of avidin is not hampered by antiavidin antibodies in humans. *Cancer Biother Radiopharm* 25, 563-570.
- Pierres, A., Touchard, D., Benoliel, A.M., and Bongrand, P. (2002). Dissecting streptavidin-biotin interaction with a laminar flow chamber. *Biophys J* 82, 3214-3223.
- Pincet, F., and Husson, J. (2005). The solution to the streptavidin-biotin paradox: the influence of history on the strength of single molecular bonds. *Biophys J* 89, 4374-4381.
- Prescher, J.A., and Bertozzi, C.R. (2005). Chemistry in living systems. *Nat Chem Biol* 1, 13-21.
- Prizant, M., Eisenberg-Domovich, Y., Hytonen, V.P., Kulomaa, M.S., Wilchek, M., Bayer, E.A., and Livnah, O. (2006). Factors dictating the pseudocatalytic efficiency of avidins. *J Mol Biol* 358, 754-763.
- Promega (2009). Technical Manual: HaloTag Protein Purification System.
- Pugliese, L., Coda, A., Malcovati, M., and Bolognesi, M. (1993). Three-dimensional structure of the tetragonal crystal form of egg-white avidin in its functional complex with biotin at 2.7 Å resolution. *J Mol Biol* 231, 698-710.
- Qureshi, M.H., and Wong, S.L. (2002). Design, production, and characterization of a monomeric streptavidin and its application for affinity purification of biotinylated proteins. *Protein Expr Purif* 25, 409-415.
- Qureshi, M.H., Yeung, J.C., Wu, S.C., and Wong, S.L. (2001). Development and characterization of a series of soluble tetrameric and monomeric streptavidin muteins with differential biotin binding affinities. *J Biol Chem* 276, 46422-46428.
- Rader, A.J. (2009). Thermostability in rubredoxin and its relationship to mechanical rigidity. *Phys Biol* 7, 16002.

- Rand, K.D., Zehl, M., Jensen, O.N., and Jorgensen, T.J. (2009). Protein hydrogen exchange measured at single-residue resolution by electron transfer dissociation mass spectrometry. *Anal Chem* **81**, 5577-5584.
- Renart, J., Reiser, J., and Stark, G.R. (1979). Transfer of proteins from gels to diazobenzyloxymethyl-paper and detection with antisera: a method for studying antibody specificity and antigen structure. *Proc Natl Acad Sci U S A* **76**, 3116-3120.
- Repo, S., Paldanius, T.A., Hytonen, V.P., Nyholm, T.K., Halling, K.K., Huuskonen, J., Pentikainen, O.T., Rissanen, K., Slotte, J.P., Airene, T.T., *et al.* (2006). Binding properties of HABA-type azo derivatives to avidin and avidin-related protein 4. *Chem Biol* **13**, 1029-1039.
- Reznik, G.O., Vajda, S., Sano, T., and Cantor, C.R. (1998). A streptavidin mutant with altered ligand-binding specificity. *Proc Natl Acad Sci U S A* **95**, 13525-13530.
- Reznik, G.O., Vajda, S., Smith, C.L., Cantor, C.R., and Sano, T. (1996). Streptavidins with intersubunit crosslinks have enhanced stability. *Nat Biotechnol* **14**, 1007-1011.
- Rico, F., and Moy, V.T. (2007). Energy landscape roughness of the streptavidin-biotin interaction. *J Mol Recognit* **20**, 495-501.
- Ringrose, A.T., Norris, L.C., and Heuser, G.F. (1931). The occurrence of a pellagra-like syndrome in chicks. *Poult Sci* **10**, 166-177.
- Romero, P.A., and Arnold, F.H. (2009). Exploring protein fitness landscapes by directed evolution. *Nat Rev Mol Cell Biol* **10**, 866-876.
- Rostovtsev, V.V., Green, L.G., Fokin, V.V., and Sharpless, K.B. (2002). A stepwise Huisgen cycloaddition process: copper(I)-catalyzed regioselective "ligation" of azides and terminal alkynes. *Angew Chem Int Ed Engl* **41**, 2596-2599.
- Saleh, O.A., Peral, C., Barre, F.X., and Allemand, J.F. (2004). Fast, DNA-sequence independent translocation by FtsK in a single-molecule experiment. *EMBO J* **23**, 2430-2439.
- Samols, D., Thornton, C.G., Murtif, V.L., Kumar, G.K., Haase, F.C., and Wood, H.G. (1988). Evolutionary conservation among biotin enzymes. *J Biol Chem* **263**, 6461-6464.
- Sano, T., and Cantor, C.R. (1990). Expression of a cloned streptavidin gene in *Escherichia coli*. *Proc Natl Acad Sci U S A* **87**, 142-146.
- Sano, T., and Cantor, C.R. (1995). Intersubunit contacts made by tryptophan 120 with biotin are essential for both strong biotin binding and biotin-induced tighter subunit association of streptavidin. *Proc Natl Acad Sci U S A* **92**, 3180-3184.

- Sano, T., Pandori, M.W., Chen, X., Smith, C.L., and Cantor, C.R. (1995). Recombinant core streptavidins. A minimum-sized core streptavidin has enhanced structural stability and higher accessibility to biotinylated macromolecules. *J Biol Chem* 270, 28204-28209.
- Sardo, A., Wohlschlager, T., Lo, C., Zoller, H., Ward, T.R., and Creus, M. (2011). Burkavidin: a novel secreted biotin-binding protein from the human pathogen *Burkholderia pseudomallei*. *Protein Expr Purif* 77, 131-139.
- Sasso, S.P., Gilli, R.M., Sari, J.C., Rimet, O.S., and Briand, C.M. (1994). Thermodynamic study of dihydrofolate reductase inhibitor selectivity. *Biochim Biophys Acta* 1207, 74-79.
- Schmidt, T.G., Koepke, J., Frank, R., and Skerra, A. (1996). Molecular interaction between the Strep-tag affinity peptide and its cognate target, streptavidin. *J Mol Biol* 255, 753-766.
- Schmidt, T.G., and Skerra, A. (1993). The random peptide library-assisted engineering of a C-terminal affinity peptide, useful for the detection and purification of a functional Ig Fv fragment. *Protein Eng* 6, 109-122.
- Schmidt-Dannert, C., and Arnold, F.H. (1999). Directed evolution of industrial enzymes. *Trends Biotechnol* 17, 135-136.
- Schrodinger, LLC (2010). The PyMOL Molecular Graphics System, Version 1.3r1.
- Seward, H.E., and Bagshaw, C.R. (2009). The photochemistry of fluorescent proteins: implications for their biological applications. *Chem Soc Rev* 38, 2842-2851.
- Shimamoto, N., Kasciukovich, T., Nagai, H., and Hayward, R.S. (1998). Efficient solubilization of proteins overproduced as inclusion bodies by use of an extreme concentration of glycerol. *J Tech Tips Online* 23, t01576.
- Shimotohno, A., Oue, S., Yano, T., Kuramitsu, S., and Kagamiyama, H. (2001). Demonstration of the importance and usefulness of manipulating non-active-site residues in protein design. *J Biochem* 129, 943-948.
- Sivanathan, V., Allen, M.D., de Bekker, C., Baker, R., Arciszewska, L.K., Freund, S.M., Bycroft, M., Lowe, J., and Sherratt, D.J. (2006). The FtsK gamma domain directs oriented DNA translocation by interacting with KOPS. *Nat Struct Mol Biol* 13, 965-972.
- Sletten, E.M., and Bertozzi, C.R. (2009). Bioorthogonal chemistry: fishing for selectivity in a sea of functionality. *Angew Chem Int Ed Engl* 48, 6974-6998.
- Sorensen, H.P., Sperling-Petersen, H.U., and Mortensen, K.K. (2003). A favorable solubility partner for the recombinant expression of streptavidin. *Protein Expr Purif* 32, 252-259.

- Stayton, P.S., Nelson, K.E., McDevitt, T.C., Bulmus, V., Shimoboji, T., Ding, Z., and Hoffman, A.S. (1999). Smart and biofunctional streptavidin. *Biomol Eng* **16**, 93-99.
- Stoddart, D., Heron, A.J., Mikhailova, E., Maglia, G., and Bayley, H. (2009). Single-nucleotide discrimination in immobilized DNA oligonucleotides with a biological nanopore. *Proc Natl Acad Sci U S A* **106**, 7702-7707.
- Suci, P.A., Kang, S., Young, M., and Douglas, T. (2009). A streptavidin-protein cage Janus particle for polarized targeting and modular functionalization. *J Am Chem Soc* **131**, 9164-9165.
- Swift, J.L., Heuff, R., and Cramb, D.T. (2006). A two-photon excitation fluorescence cross-correlation assay for a model ligand-receptor binding system using quantum dots. *Biophys J* **90**, 1396-1410.
- Takakura, Y., Tsunashima, M., Suzuki, J., Usami, S., Kakuta, Y., Okino, N., Ito, M., and Yamamoto, T. (2009). Tamavidins--novel avidin-like biotin-binding proteins from the Tamogitake mushroom. *FEBS J* **276**, 1383-1397.
- Taki, M., Hohsaka, T., Murakami, H., Taira, K., and Sisido, M. (2001). A non-natural amino acid for efficient incorporation into proteins as a sensitive fluorescent probe. *FEBS Lett* **507**, 35-38.
- Tausig, F., and Wolf, F.J. (1964). Streptavidin--a substance with avidin-like properties produced by microorganisms. *Biochem Biophys Res Commun* **14**, 205-209.
- Terwilliger, T.C., Grosse-Kunstleve, R.W., Afonine, P.V., Moriarty, N.W., Zwart, P.H., Hung, L.W., Read, R.J., and Adams, P.D. (2008). Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr D Biol Crystallogr* **64**, 61-69.
- Thoma, R.W., and Peterson, W.H. (1954). The enzymatic degradation of soluble bound biotin. *J Biol Chem* **210**, 569-579.
- Thompson, L.D., and Weber, P.C. (1993). Construction and expression of a synthetic streptavidin-encoding gene in *Escherichia coli*. *Gene* **136**, 243-246.
- Thompson, M.J., and Eisenberg, D. (1999). Transproteomic evidence of a loop-deletion mechanism for enhancing protein thermostability. *J Mol Biol* **290**, 595-604.
- Thomson, A.M., and Armstrong, W.E. (2011). Biocytin-labelling and its impact on late 20th century studies of cortical circuitry. *Brain Res Rev* **66**, 43-53.
- Tokuda, Y., Watanabe, T., Horiike, K., Shiraga, K., Abe, R., Muranaka, N., and Hohsaka, T. (2011). Biosynthesis of proteins containing modified lysines and fluorescent labels using non-natural amino acid mutagenesis. *J Biosci Bioeng* **111**, 402-407.

- Tronrud, D.E. (1996). Knowledge-Based B-Factor Restraints for the Refinement of Proteins. *Journal of Applied Crystallography* 29, 100-104.
- Tsien, R.Y. (1998). The green fluorescent protein. *Annu Rev Biochem* 67, 509-544.
- Turner, P., Holst, O., and Karlsson, E.N. (2005). Optimized expression of soluble cyclomaltodextrinase of thermophilic origin in *Escherichia coli* by using a soluble fusion-tag and by tuning of inducer concentration. *Protein Expr Purif* 39, 54-60.
- Umetsu, M., Tsumoto, K., Hara, M., Ashish, K., Goda, S., Adschiri, T., and Kumagai, I. (2003). How additives influence the refolding of immunoglobulin-folded proteins in a stepwise dialysis system. Spectroscopic evidence for highly efficient refolding of a single-chain Fv fragment. *J Biol Chem* 278, 8979-8987.
- Unciuleac, M.C., and Shuman, S. (2010). Characterization of the mycobacterial AdnAB DNA motor provides insights into the evolution of bacterial motor-nuclease machines. *J Biol Chem* 285, 2632-2641.
- Vallejo, L.F., and Rinas, U. (2004). Strategies for the recovery of active proteins through refolding of bacterial inclusion body proteins. *Microb Cell Fact* 3, 11.
- Vincentelli, R., Canaan, S., Campanacci, V., Valencia, C., Maurin, D., Frassinetti, F., Scappucini-Calvo, L., Bourne, Y., Cambillau, C., and Bignon, C. (2004). High-throughput automated refolding screening of inclusion bodies. *Protein Sci* 13, 2782-2792.
- Visnapuu, M.L., and Greene, E.C. (2009). Single-molecule imaging of DNA curtains reveals intrinsic energy landscapes for nucleosome deposition. *Nat Struct Mol Biol* 16, 1056-1062.
- Voigt, N.V., Topping, T., Rotaru, A., Jacobsen, M.F., Ravnsbaek, J.B., Subramani, R., Mamdouh, W., Kjems, J., Mokhir, A., Besenbacher, F., *et al.* (2010). Single-molecule chemical reactions on DNA origami. *Nat Nanotechnol* 5, 200-203.
- Voss, S., and Skerra, A. (1997). Mutagenesis of a flexible loop in streptavidin leads to higher affinity for the Strep-tag II peptide and improved performance in recombinant protein purification. *Protein Eng* 10, 975-982.
- Wang, L., and Lutkenhaus, J. (1998). FtsK is an essential cell division protein that is localized to the septum and induced as part of the SOS response. *Mol Microbiol* 29, 731-740.
- Wang, W.W., Das, D., and Suresh, M.R. (2005). Biotin carboxyl carrier protein co-purifies as a contaminant in core-streptavidin preparations. *Mol Biotechnol* 31, 29-40.

- Weber, P.C., Ohlendorf, D.H., Wendoloski, J.J., and Salemme, F.R. (1989). Structural origins of high-affinity biotin binding to streptavidin. *Science* **243**, 85-88.
- Weber, P.C., Pantoliano, M.W., Simons, D.M., and Salemme, F.R. (1994). Structure-Based Design of Synthetic Azobenzene Ligands for Streptavidin. *Journal of the American Chemical Society* **116**, 2717-2724.
- Weber, P.C., Wendoloski, J.J., Pantoliano, M.W., and Salemme, F.R. (1992). Crystallographic and Thermodynamic Comparison of Natural and Synthetic Ligands Bound to Streptavidin. *Journal of the American Chemical Society* **114**, 3197-3200.
- Weiden, P.L., Breitz, H.B., Press, O., Appelbaum, J.W., Bryan, J.K., Gaffigan, S., Stone, D., Axworthy, D., Fisher, D., and Reno, J. (2000). Pretargeted radioimmunotherapy (PRIT) for treatment of non-Hodgkin's lymphoma (NHL): initial phase I/II study results. *Cancer Biother Radiopharm* **15**, 15-29.
- Weiss, G.A., Watanabe, C.K., Zhong, A., Goddard, A., and Sidhu, S.S. (2000). Rapid mapping of protein functional epitopes by combinatorial alanine scanning. *Proc Natl Acad Sci U S A* **97**, 8950-8954.
- Wilbur, D.S., Hamlin, D.K., Buhler, K.R., Pathare, P.M., Vessella, R.L., Stayton, P.S., and To, R. (1998). Streptavidin in antibody pretargeting. 2. Evaluation of methods for decreasing localization of streptavidin to kidney while retaining its tumor binding capacity. *Bioconjug Chem* **9**, 322-330.
- Wilbur, D.S., Hamlin, D.K., Meyer, D.L., Mallett, R.W., Quinn, J., Vessella, R.L., and Press, O.W. (2002). Streptavidin in antibody pretargeting. 3. Comparison of biotin binding and tissue localization of 1,2-cyclohexanedione and succinic anhydride modified recombinant streptavidin. *Bioconjug Chem* **13**, 611-620.
- Wilbur, D.S., Hamlin, D.K., Sanderson, J., and Lin, Y. (2004). Streptavidin in antibody pretargeting. 4. Site-directed mutation provides evidence that both arginine and lysine residues are involved in kidney localization. *Bioconjug Chem* **15**, 1454-1463.
- Wilchek, M., and Bayer, E.A. (1999). Foreword and introduction to the book (strept)avidin-biotin system. *Biomol Eng* **16**, 1-4.
- Williams, D.H., Stephens, E., and Zhou, M. (2003). Ligand binding energy and catalytic efficiency from improved packing within receptors and enzymes. *J Mol Biol* **329**, 389-399.
- Williams, J.G., Steffens, D.L., Anderson, J.P., Urlacher, T.M., Lamb, D.T., Grone, D.L., and Egelhoff, J.C. (2008). An artificial processivity clamp made with streptavidin facilitates oriented attachment of polymerase-DNA complexes to surfaces. *Nucleic Acids Res* **36**, e121.

- Wolbers, F., ter Braak, P., Le Gac, S., Luttge, R., Andersson, H., Vermes, I., and van den Berg, A. (2006). Viability study of HL60 cells in contact with commonly used microchip materials. *Electrophoresis* 27, 5073-5080.
- Wolf, B., Grier, R.E., Secor McVoy, J.R., and Heard, G.S. (1985). Biotinidase deficiency: a novel vitamin recycling defect. *J Inherit Metab Dis* 8 *Suppl* 1, 53-58.
- Wray, J.W., Baase, W.A., Lindstrom, J.D., Weaver, L.H., Poteete, A.R., and Matthews, B.W. (1999). Structural analysis of a non-contiguous second-site revertant in T4 lysozyme shows that increasing the rigidity of a protein can enhance its stability. *J Mol Biol* 292, 1111-1120.
- Wu, S.C., and Wong, S.L. (2005). Engineering soluble monomeric streptavidin with reversible biotin binding capability. *J Biol Chem* 280, 23225-23231.
- Wu, S.C., and Wong, S.L. (2006). Intracellular production of a soluble and functional monomeric streptavidin in *Escherichia coli* and its application for affinity purification of biotinylated proteins. *Protein Expr Purif* 46, 268-273.
- Yan, H., Park, S.H., Finkelstein, G., Reif, J.H., and LaBean, T.H. (2003). DNA-templated self-assembly of protein arrays and highly conductive nanowires. *Science* 301, 1882-1884.
- Yates, J., Zhekov, I., Baker, R., Eklund, B., Sherratt, D.J., and Arciszewska, L.K. (2006). Dissection of a functional interaction between the DNA translocase, FtsK, and the XerD recombinase. *Mol Microbiol* 59, 1754-1766.
- Ye, J., Osborne, A.R., Groll, M., and Rapoport, T.A. (2004). RecA-like motor ATPases--lessons from structures. *Biochim Biophys Acta* 1659, 1-18.
- Ying, L.Q., and Branchaud, B.P. (2011). Design of a reversible biotin analog and applications in protein labeling, detection, and isolation. *Chem Commun (Camb)* 47, 8593-8595.
- Yuan, C., Chen, A., Kolb, P., and Moy, V.T. (2000). Energy landscape of streptavidin-biotin complexes measured by atomic force microscopy. *Biochemistry* 39, 10219-10223.
- Yuan, L., Kurek, I., English, J., and Keenan, R. (2005a). Laboratory-directed protein evolution. *Microbiol Mol Biol Rev* 69, 373-392.
- Yuan, Z., Bailey, T.L., and Teasdale, R.D. (2005b). Prediction of protein B-factor profiles. *Proteins* 58, 905-912.
- Zempleni, J. (2005). Uptake, localization, and noncarboxylase roles of biotin. *Annu Rev Nutr* 25, 175-196.
- Zhang, Y., Olsen, D.R., Nguyen, K.B., Olson, P.S., Rhodes, E.T., and Mascarenhas, D. (1998). Expression of eukaryotic proteins in soluble form in *Escherichia coli*. *Protein Expr Purif* 12, 159-165.

Zhu, D., Tang, Y., Xing, D., and Chen, W.R. (2008). PCR-free quantitative detection of genetically modified organism from raw materials. An electrochemiluminescence-based bio bar code method. *Anal Chem* 80, 3566-3571.

Publications

The article originally located here has been removed from this version of the thesis for copyright reasons. Article details: Chivers, C. E. et al. (2010). 'A streptavidin variant with slower biotin dissociation and increased mechanostability', *Nature Methods* 7(5), 391-396. Available at: <http://dx.doi.org/10.1338/nmeth.1450>

The article originally located here has been removed from this version of the thesis for copyright reasons. Article details: Chivers, C. E. et al. (2010). 'A streptavidin variant with slower biotin dissociation and increased mechanostability', *Nature Methods* 7(5), 391-396. Available at: <http://dx.doi.org/10.1338/nmeth.1450>

The article originally located here has been removed from this version of the thesis for copyright reasons. Article details: Chivers, C. E. et al. (2010). 'A streptavidin variant with slower biotin dissociation and increased mechanostability', *Nature Methods* 7(5), 391-396. Available at: <http://dx.doi.org/10.1338/nmeth.1450>

Separating speed and ability to displace roadblocks during DNA translocation by FtsK

This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits distribution, and reproduction in any medium, provided the original author and source are credited. This license does not permit commercial exploitation without specific permission.

Estelle Crozat¹, Adrien Meglio²,
Jean-François Allemand², Claire E Chivers¹,
Mark Howarth¹, Catherine Vénien-Bryan¹,
Ian Grainge^{1,3} and David J Sherratt^{1,*}

¹Department of Biochemistry, University of Oxford, Oxford, UK
and ²Laboratoire de Physique Statistique et Département de Biologie,
Ecole Normale Supérieure, UPMC, Paris 06, Université Paris Diderot,
CNRS, Paris, France

FtsK translocates dsDNA directionally at >5 kb/s, even under strong forces. *In vivo*, the action of FtsK at the bacterial division septum is required to complete the final stages of chromosome unlinking and segregation. Despite the availability of translocase structures, the mechanism by which ATP hydrolysis is coupled to DNA translocation is not understood. Here, we use covalently linked translocase subunits to gain insight into the DNA translocation mechanism. Covalent trimers of wild-type subunits dimerized efficiently to form hexamers with high translocation activity and an ability to activate XerCD-*dif* chromosome unlinking. Covalent trimers with a catalytic mutation in the central subunit formed hexamers with two mutated subunits that had robust ATPase activity. They showed wild-type translocation velocity in single-molecule experiments, activated translocation-dependent chromosome unlinking, but had an impaired ability to displace either a triplex oligonucleotide, or streptavidin linked to biotin-DNA, during translocation along DNA. This separation of translocation velocity and ability to displace roadblocks is more consistent with a sequential escort mechanism than stochastic, hand-off, or concerted mechanisms.

The EMBO Journal (2010) 29, 1423–1433. doi:10.1038/

emboj.2010.29; Published online 8 April 2010

Subject Categories: genome stability & dynamics

Keywords: *Escherichia coli* chromosome segregation; FtsK
DNA translocase; molecular motor

Introduction

FtsK is a highly conserved dsDNA translocase that functions in the final stages of bacterial chromosome unlinking and in coordinating this final stage of chromosome segregation with

cell division (Begg *et al*, 1995; Yu *et al*, 1998; Recchia *et al*, 1999; Aussel *et al*, 2002; Grainge *et al*, 2007). FtsK is tethered to the developing division septum by its membrane-spanning N-terminal domain and targets its activity to any dimeric, catenated or incompletely replicated chromosomes in the G2 phase of the cell cycle (Steiner *et al*, 1999; Dorazi and Dewar, 2000; Aussel *et al*, 2002).

The C-terminal ~500 amino-acid (aa) FtsK translocase is composed of three subdomains: $\alpha\beta$ constitute the active motor, whereas the γ -subdomain has regulatory functions. The translocase loads FtsK preferentially at specific DNA sequences that are oriented in the bacterial chromosome. In *Escherichia coli*, FtsK Oriented Polar Sequences (KOPS; GGGNAGGG) are the preferred loading sites from where FtsK translocates DNA at ~5 kb/s (Saleh *et al*, 2004; Pease *et al*, 2005; Bigot *et al*, 2006). Finally, the translocation by FtsK stops at XerCD-bound *dif* sites in the replication terminus region, where FtsK activates the final chromosome unlinking by XerCD recombination at *dif* (Graham *et al*, 2010). The γ -regulatory subdomain of FtsK acts in KOPS loading and in the activation of XerCD recombination at *dif*.

The FtsK translocase belongs to the RecA subgroup of the additional strand catalytic glutamate (ASCE) ATPase superfamily that also contain AAA+ (ATPases associated with various cellular activities) enzymes that have multiple biological roles, including translocating along nucleic acids in the 3'–5' direction. In contrast, translocases and helicases that belong to the RecA group (e.g. T7gp4 and DnaB), translocate in the 5'–3' direction in the cases analysed (Singleton *et al*, 2007; Thomsen and Berger, 2009). It, therefore, seems likely that during FtsK dsDNA translocation, the catalytic mechanism uses just one of the two DNA chains and that translocation is 5'–3' with respect to this chain. The nature of the ASCE folds places the ATP-binding pocket of one subunit close to γ -phosphate sensor elements in an adjacent subunit, thereby providing the opportunity for coordinated catalysis and ATP hydrolysis. A crystal structure of the $\alpha\beta$ motor of *Pseudomonas aeruginosa* FtsK bound to ADP showed a symmetrical hexamer with no obvious loops protruding into the central chain in which ds DNA is normally located (Massey *et al*, 2006). This led to the proposal of a sequential rotary mechanism for translocation, with a rotational resetting after each catalytic step, which was proposed to translocate 2 bp of DNA. Nevertheless, the structure did not provide any evidence for the asymmetry in structures and nucleotide-binding states that one would predict from such a mechanism.

To further address the FtsK translocation mechanism, we have covalently linked the minimal predicted functional translocase, identified by the crystallographic studies, into multimers using a 14 aa flexible linker. This strategy was previously used to study the hexameric ClpX protein-unfolding motor (Martin *et al*, 2005) and the recombination protein

*Corresponding author. Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK. Tel.: +44 1865 613 237; Fax: +44 1865 613 238; E-mail: david.sherratt@bioch.ox.ac.uk

³Present address: BG09, Biological Sciences, Faculty of Science and Information Technology, University of Newcastle, Callaghan, NSW, Australia.

Received: 18 December 2009; accepted: 12 February 2010; published online: 8 April 2010

RecA (Chen *et al*, 2008). Intriguingly, some AAA + motors have their six AAA + domains naturally covalently linked into a single peptide (Vallee and Hook, 2006). This strategy allows us to introduce mutations into defined subunits and study the translocation properties of the mutated protein to understand how nucleotide binding and hydrolysis are coupled to translocation and the ability to do mechanical work. We demonstrate that covalent trimers dimerize to form hexamers that recognize KOPS, activate chromosome unlinking and translocate at velocities comparable to FtsK_{50C}, the previously used translocase. Furthermore, we show that trimers with a catalytically inactive subunit at the central position form hexamers that have a normal translocation velocity but which are impaired in their ability to displace 'roadblocks' on DNA.

Results

FtsK translocase multimers are active *in vivo*

Most previous research on FtsK translocation *in vitro* has used a protein, FtsK_{50C}, in which a 50 aa segment derived from the FtsK N-terminus has been added to the $\alpha\beta\gamma$ translocase domain (Aussel *et al*, 2002; Ip *et al*, 2003; Saleh *et al*, 2004; Pease *et al*, 2005; Bigot *et al*, 2006). These 50 aa facilitate hexamerization and are required for significant *in vitro* catalytic activity (Aussel *et al*, 2002). Nevertheless, FtsK_{50C} has a high propensity to aggregate, making quantitative biochemistry and mechanistic interpretation difficult. We reasoned that a minimal FtsK translocase, lacking the N-terminal 50 aa of FtsK_{50C} and derived by using the available structure (Massey *et al*, 2006), might demonstrate *in vitro* activity if the individual monomers were covalently linked. Therefore, FtsK aa 840–1329 were linked together with a 14 aa linker (GGGSEGGGSEGGSG), thus forming covalent multimers of the translocase, the first subunit being tagged with a 6-His tag and/or biotin-tagged peptide (Figure 1).

In initial experiments, we showed that the gene, encoding a 320 kDa FtsK covalent hexamer, expressed well enough to give high levels of *in vivo* translocation-dependent XerCD-*dif* recombination (data not shown). Nevertheless, the level of expression was not sufficient to enable ready purification of the protein. Therefore, we chose to work with monomers, covalent dimers and covalent trimers (hereafter described as 'trimers'), which were expressed well. This provided the opportunity to introduce specific mutations into defined individual subunits of a covalent multimer. We then tested the *in vivo* activity of these translocases.

First, we exploited the observation that the overexpression of active FtsK translocase (e.g. FtsK_{50C} when expressed from a *para* promoter with 0.2% arabinose) is toxic and cells stop growing and die within ~15 min of induced expression. Toxicity is correlated with the ability to hydrolyse ATP (Massey *et al*, 2006). Expression of monomer, dimer and trimer of the new FtsK translocase derivatives was toxic, consistent with them having *in vivo* activity (Table I). Cells stopped growing ~45 min after induced expression of the monomer, ~30 min after dimer expression and ~15 min after trimer expression, consistent with multimerization by covalent linking enhancing specific activity, because hexamers are the active species. Second, we tested whether the proteins could support *in vivo* XerCD-*dif* recombination on a reporter plasmid containing two *dif* sites (Recchia *et al*,

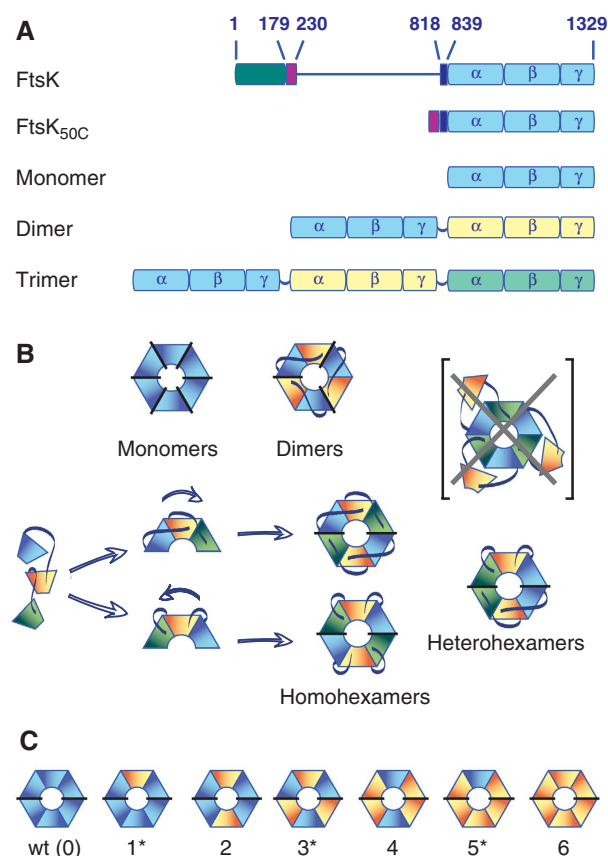


Figure 1 (A) Schematic of the FtsK proteins used. FtsK depicts the wild-type protein, with four transmembrane helices in the N-terminus region (dark green), a 639 amino-acid linker (blue line) and the C-terminus motor domain (light blue boxes) containing the three subdomains α , β and γ . FtsK_{50C} is the soluble *E. coli* FtsK derivative that has been used in previous *in vitro* studies. The pink box represents the 50 aa segment derived from the linker, which is directly linked to the C-terminal motor. The dark blue box is a segment absent in the FtsK_{50C} structure and in the derivatives used here: monomer, covalent dimers and covalent trimers. Subunits in covalent multimer derivatives are connected by a 14 aa linker, joining directly the C-terminus of the γ -subdomain to the first amino acid of the motor, L840. (B) Different potential configurations for multimer formation. The 14 aa linker is sufficiently short that we expect subunit 1 (blue) to be always adjacent to 2 (yellow), and subunit 3 (green) adjacent to 2 in the trimers. Nevertheless, it is plausible that two types of trimers can form, with the subunits folding in either a clockwise, or anticlockwise sequence. These can then potentially form mixed hexamers (heterohexamers) or unmixed hexamers (homohexamers). Because of the uncertainty of the configuration in trimers, mutations were introduced into either subunit 2 (single mutants), or subunits 1 and 3 (double mutants). In parentheses, with a cross, is shown how three covalent trimers with a centrally placed mutated subunit could conceivably form a hexamer with wild-type subunits and three looped out mutated subunits. We have no evidence that this can form. (C) Mutant hexamers. This figure depicts the mutants used in this study. Pure hexamers are obtained with a wild-type, single-mutant trimer, double-mutant trimer or triple-mutant trimer (0, 2, 4 or 6). Mixes of trimers are required to form hexamers with 1, 3 or 5 mutant subunits (marked with an asterisk). Mutated subunits are yellow and wild-type subunits are blue.

1999). The results of this assay mirrored the toxicity results; trimers were more active than covalent dimers, which were more active than monomers (Table I). These semi-quantitative assays do not take into account any differential levels of

Table 1 *In vivo* activities of FtsK multimers

Assay	Monomer	Dimer	wt Trimer	2 WA ^a	4 WA	6 WA	2 WB	4 WB	6 WB
Toxicity ^b (min)	45	30	15	25	30	—	25	35	—
Recombination ^c (%)	25	50	60	63	12	17	58	24	24

^aHexamers formed by dimerization of trimers are described by the number of mutated WA or WB subunits.

^bToxicity was assayed as the time at which cell growth in LB ceases after FtsK induction with 0.2% arabinose. '—' indicates no toxicity.

^cRecombination was measured *in vivo* on a reporter plasmid, 2 h after induction of FtsK expression by 0.02% arabinose. The values shown have the 0 min background subtracted.

these proteins in cells. Nevertheless, our extensive experience with these assays using FtsK_{50C} derivatives (Sivanathan *et al*, 2006 and data not shown) has shown them to give a good indication of *in vitro* specific activity. We conclude that covalent trimers and dimers are highly active *in vivo*.

***In vitro* activities of FtsK translocase multimers**

The recombinant FtsK monomer, covalent dimer and trimer were then purified and assayed *in vitro* for ATPase activity and for two independent assays of DNA translocation: *in vitro* FtsK-dependent XerCD-*dif* recombination and triplex displacement assays (Figure 2; Aussel *et al*, 2002; Sivanathan *et al*, 2006). ATPase activity was DNA dependent (not shown), and at a concentration of 50 nM hexamer, the trimer hydrolysed $>1.7 \times 10^3$ ATP/min/hexamer. At this protein concentration, the covalent dimers displayed ~80% of trimer specific activity, and the monomers ~1%, because of less efficient formation of active hexamers. The relatively low steady state level of ATP hydrolysis by the trimers indicates that most molecules are not translocating at any given time, presumably because they are not loaded onto DNA, have collided with other translocases, or are simply not active.

In XerCD-*dif* recombination and triplex displacement assays, the trimers were again the most active species, with covalent dimers showing ~40% of the activity of trimers. In the triplex displacement assays, optimal activity was obtained with trimers at 50 nM (hexamer), but required 80 and 150 nM for covalent dimers and monomers, respectively. These results confirm that covalent dimers and trimers are more active *in vitro* than the constituent monomers, with dimers and monomers forming hexamers at high and very high concentrations, respectively. In conclusion, trimers dimerize to form active hexamers efficiently at 50 nM, and their activity is higher than that observed for FtsK_{50C} at the same concentration (Figure 1C and data not shown). Furthermore, the trimers were still responsive to the presence of KOPS in the triplex displacement and recombination assays (data not shown).

FtsK translocase trimers form hexamers on DNA

Despite the trimers being highly active in translocation, we needed to be confident that the activity was resulting from dimerization of two trimers into a hexamer, rather than the activity residing in higher-order structures (Figure 1B, right panel), which could compromise interpretation of data once defined mutations had been introduced into the linked multimers. FtsK subunits were rendered non-functional by mutation in the highly conserved ATP-binding pocket. We used two mutations that have been shown to be catalytically defective in an extensive range of studies of RecA-fold proteins (ClpB: Watanabe *et al*, 2002; Dynein: Reck-Peterson and Vale, 2004; Rad51: Wiese *et al*, 2006 and archeal MCM:

Moreau *et al*, 2007). A K997A substitution in the Walker A box (WA) should prevent ATP binding, whereas a D1121A substitution in the Walker B box (WB) may allow ATP binding, but should be defective in ATP hydrolysis. Binding of trimers, containing 0 to 3 WA mutated subunits, to short KOPS-containing DNA was analysed by gel electrophoresis. Two shifted bands were observed in the presence of 200 nM protein, consistent with binding of one and two trimers on DNA (data not shown). Protein-DNA complexes containing wild-type, or three mutated WA subunits, were also observed by electron microscopy using negative staining. On a 2.7 kb DNA, particles whose shape and size were consistent with 'side-view' FtsK hexamers were the majority species (Figure 2E; Massey *et al*, 2006). The trapezium shape particles had diameters of 120–130 Å (large side) and 60–70 Å (small side). Aggregates or higher-order structures comprised <2% of particles. On a short DNA (44 bp), when 'end-on' view (top view) particles are expected, rounded particles of diameter ~130 Å were observed (Figure 2E), the size expected for a hexamer obtained by dimerization of trimers. However, we did not observe any lack of density corresponding to the central hole, or any six-fold symmetry (Massey *et al*, 2006). This may be because the γ -subdomains are attached to the $\alpha\beta$ motor by a flexible linker, and therefore may occlude the central hole, as well as mask the symmetry of the protein complexes.

We then analysed whether the trimers could adopt alternative configurations when they dimerize into hexamers. The 14 aa linker may be long enough that it does not restrict the assembly into only a single configuration, although we expect the linker length to restrict the arrangement of subunits such that subunit 1 is never adjacent to subunit 3 (Figure 1B). The existence of alternative subunit configurations potentially complicates the interpretation of experiments using mutant subunits. To address this issue, we labelled His-tagged covalent trimers, pre-loaded onto short DNA to give preferential end-views, with 2 nm Ni-NTA-bound gold beads and assessed the relative locations of the beads within the hexamers by electron microscopy, followed by image classification (Figure 2G). The 1702 particles analysed were classified into 10, 50 or 100 groups on the basis of protein shape and gold position. Classes with two adjacent golds and with the two golds *in trans* were observed for each group, indicating that trimers can dimerize in both head-tail and head-head configurations (Figures 1B and 2G). However, some classes were ambiguous, with a very low signal-noise ratio, thereby precluding quantitative estimates of the two configurations. Given the likelihood that both configurations can form, we chose symmetrical configurations of mutations to avoid any ambiguity in the configurations of the subunits in the assembled hexamers. We focused most of our experiments on trimers containing a mutated subunit in the central

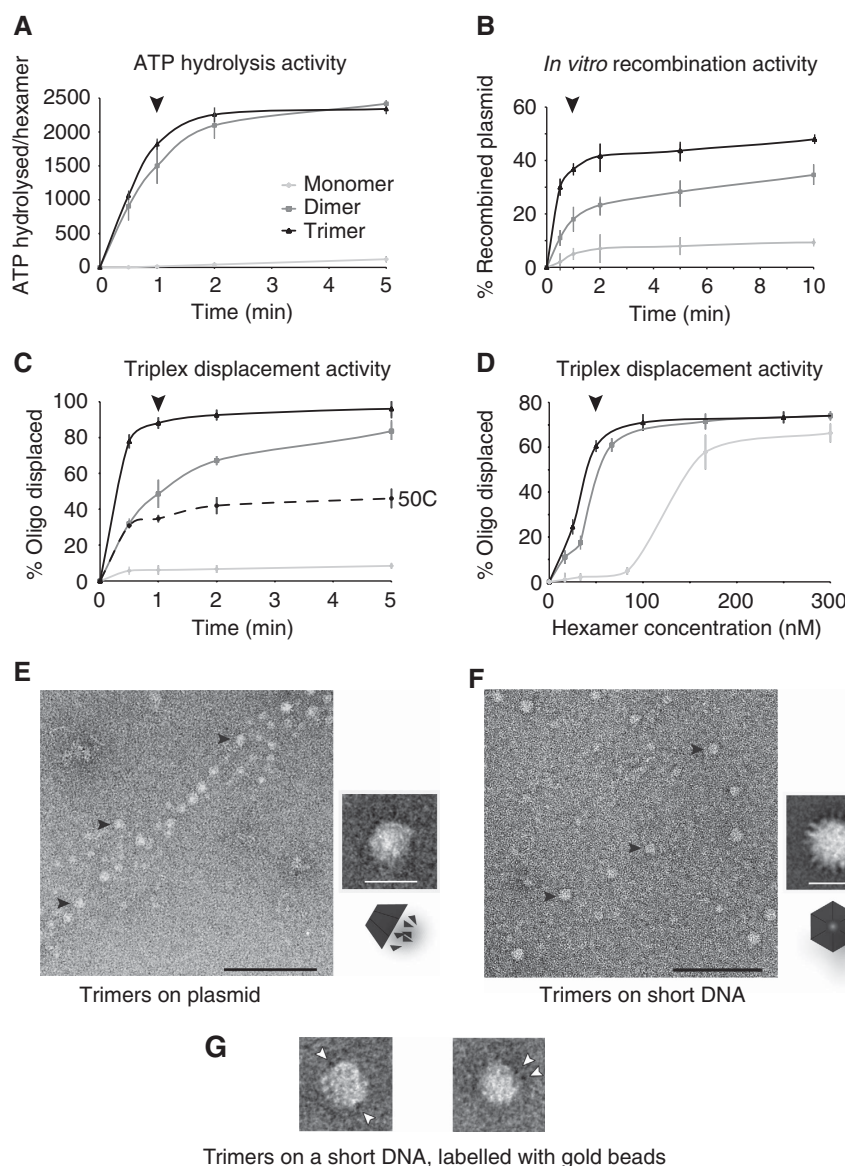


Figure 2 *In vitro* activity of FtsK multimers. **(A)** ATP hydrolysis over time is shown for monomer, covalent dimer and trimer, at a concentration of 50 nM hexamer equivalent. The black arrow indicates the 1 min time point, which was used in later experiments (Figure 3). The mean data from three independent experiments, with standard deviations, are shown in panels **(A–D)**. **(B)** FtsK-dependent XerCD-*dif* recombination. Recombination was followed over time on a dimeric plasmid containing two *dif* sites lacking consensus KOPS-loading sites. Proteins (50 nM hexamer equivalent) were incubated with ATP and pre-bound XerCD, and the 1 min time point data (arrow) were used to generate the data in Figure 3. **(C)** Triplex displacement as a function of time; protein concentration was 50 nM (hexamer equivalent); the dotted line shows FtsK_{50C} activity as a comparison. **(D)** Triplex displacement as a function of protein concentration; 1 min reactions were used to compare translocation activity of monomer, dimers and trimers over a range of 0–350 nM (hexamer equivalent). The activity does not increase with higher concentrations of protein (data not shown). **(E–G)** Electron microscopy of covalent trimers of FtsK on DNA. **(E)** A 2.7 kb, linearized plasmid, was used to study DNA binding by the trimers. A representative field using wild-type trimer is shown, as well as an averaged image of the protein seen from the side (162 particles), showing that the particles have the overall size expected for a hexamer composed of two trimers. Similar results were obtained with double and triple WA mutant trimers (data not shown). The black bar represents 100 nm, and the white bar 13 nm. Arrows point at examples of particles and a schematic of the side-on view is drawn under the averaged reconstructed image (inset). **(F)** Wild-type trimers were incubated with a 44 bp DNA-containing KOPS, thereby leading to preferential top-bottom views, giving an averaged reconstruction of particles of ~13 nm diameter (2600 particles selected) (inset). Similar results were obtained with double and triple WA mutant trimers (data not shown). Scale bars are the same as **(E)**. The schematic indicates the orientation of the particle on the grid. **(G)** Examples of gold labelling. Two classes averaged from a 100 group classification are shown, with beads on opposite sides of the particle (left) or side-by-side (right).

position, so that hexamers should carry two mutated subunits each separated by two wild-type subunits. Taken together, these results give us confidence that most biological activities assayed were the result of dimerization of trimers on DNA.

Activities of translocases with mutated subunits

To study coordination between FtsK subunits during DNA translocation, we compared the activities of presumptive hexamers derived from trimers with unmutated and/or mutated subunits. The mutations were introduced separately

into a given subunit of the trimers; in the central position to get a single mutant; in the first and the last subunits to get a double mutant and in all three subunits to get a triple mutant.

The mutated trimers were first tested *in vivo* using the toxicity and XerCD-*dif* recombination assays (Table I). In the toxicity assay, cells expressing the wild-type trimer showed a loss of exponential growth 15 min after induction of expression. Trimers containing three WA or three WB mutated subunits showed no toxicity and continued to grow. Trimers containing single WA or WB mutations were more toxic than trimers carrying two WA or WB mutations, with the trimer with two mutated WA subunits appearing slightly less toxic than its WB counterpart. The *in vivo* FtsK XerCD-*dif* recombination assays broadly mirrored the toxicity results, although in these assays, trimers with a single WA or WB mutant showed similar activity to wild-type trimers, whereas trimers with two WA or WB mutations appeared as deficient as trimers with three mutations. The residual level of recombination from the trimers with three mutated subunits may result from translocation-independent stimulation of recombination by the γ -subunits present on the trimers (Grainge I and Sherratt DJ, unpublished data).

Wild-type trimers were more active than wild-type covalent dimers in both of these assays, whereas wild-type monomers showed little if any activity. This result reassured us that any monomers or covalent dimers with one wild-type subunit and one mutated subunit, derived from proteolysis of trimers *in vitro* or *in vivo*, are unlikely to have significant activity that could confound the interpretation of the above experiments. These results gave us the confidence to purify the mutated trimers, giving the single, double and triple mutants in the WA and WB motifs, respectively (Figure 1).

The FtsK translocation mechanism is neither stochastic nor strictly concerted

FtsK-dependent XerCD-*dif* site-specific recombination and displacement of a triplex-forming oligonucleotide were used to assay DNA translocation by the trimer variants and their combinations (Figure 3). For trimer mutants with 0, 1, 2 or 3 mutated subunits, we followed the decrease of recombination as a function of the number of mutant subunits. Assuming that the only potentially active species are hexamers formed by trimer dimerization, this gives data for hexamers with 0, 2, 4, or 6 mutated subunits. To obtain results for hexamers containing 1, 3 or 5 mutant subunits, different trimer mutants were mixed in equimolar amounts (Figure 1). For example, to obtain a hexamer population with only one mutated subunit, equal amounts of wild-type trimer and single-mutant trimer were mixed. In this mixture, it is predicted that 25% of hexamers are wild type; 50% carry a single mutated subunit; and 25% carry two mutated subunits. As the activity of each homohexamer is known, we were able to deduce the activity of the heterohexamer.

For the *in vitro* recombination assays that are dependent on DNA translocation by FtsK for activation of recombination, hexamers in which all of the subunits carry WA or WB mutations showed no activity, as expected, and gave us confidence that the mutations abrogate DNA translocation. By comparison, the wild-type subunits showed ~30% recombinant product in a 1 min reaction (Figure 2B; normalized to 100% in Figure 3).

Hexamers with two WA or WB mutations showed substantial recombination activity (~65% of wild type in 1 min reactions), whereas the mixed subunit population that gives hexamers with a single mutation (50%) showed at least wild-type activities after correction for the wild-type (25%) and double-mutated (25%) population. Some residual

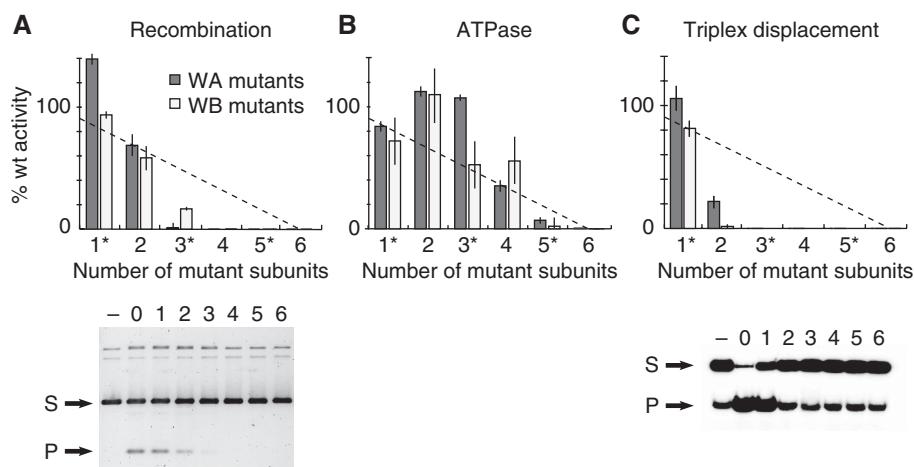


Figure 3 *In vitro* activity of covalently linked mutated hexamers. (A) XerCD-*dif* recombination as a function of the number of mutated subunits. Recombination reactions were performed for 1 min on a dimeric plasmid containing two *dif* sites. In all, 50 nM hexamers were used on 27 nM 5.6 kb plasmid. % recombination is normalized to the activity of the wild-type (wt) hexamer, which gave ~30% recombination in 1 min (Figure 2). The dotted line shows the levels if there were a linear activity decrease as a function of increasing number of mutated subunits (A–C). The panel below shows an example of a recombination gel (0.8% agarose, 1 × TAE) for hexamers containing 0–6 WB mutated subunits (S) denotes recombination substrate; (P) denotes product; (–) denotes no FtsK. Error bars indicate standard deviation in three independent experiments, and asterisks indicate the combinations made by mixing appropriate trimers as in Figure 1C (A–C). (B) ATPase activity measured on a fraction of the recombination reaction shown in (A). Activity is normalized to that of the wild-type hexamer. (C) Triplex displacement. Triplex-displacement assays were carried out as in Figure 2. After subtracting background, activity was normalized to that of the wild-type hexamer. The panel below shows an example of a triplex reaction gel (4% acrylamide, 1 × TAM) for hexamers containing 0–6 WB mutants.

recombination activity (<20%) was observed in the assays in which 50% of the population should contain hexamers with three mutated subunits. No activity was observed with four mutated subunits. These observations are inconsistent with a stochastic firing model, in which a linear decrease in activity as a function of increasing mutated subunit number would be expected (Moreau *et al*, 2007). Similarly, obligate sequential rotary models, in which each subunit must be catalytically active in turn, can be ruled out, because a single-mutated subunit would block translocation. In parallel, ATPase activity was assayed in each of the recombination reactions, giving an estimate of the number of ATP consumed as a function of the level of recombination for each mutant hexamer (Figure 3B). Note that the absolute level of ATPase activity in the recombination assays cannot be compared directly with ATPase activity measured in the absence of recombinases, because the ATPase activity is downregulated when FtsK encounters XerCD bound to *dif* (Graham *et al*, 2010). ATPase activity decreased less sharply, with the four mutated subunit population retaining >40% of wild-type activity. This observation is consistent with an uncoupling of ATPase and translocation. Indeed by independent assays of ATPase activity under optimal conditions, with DNA in excess, there was an approximately linear decrease in ATPase activity as a function of the number of mutated subunits (data not shown). These results contrast to those from experiments in which wild-type *P. aeruginosa* FtsK (PAK4) hexamers were doped with increasing amounts of a WA mutant subunit and the ATPase activity measured (Massey *et al*, 2006). In that paper, there was a rapid, greater than linear reduction in ATPase activity as a function of mutated subunit concentration, leading to the conclusion of a rigorous sequential (or concerted) mechanism. However, we note that the absolute levels of ATPase activity were low, perhaps because a 16 bp dsDNA was used, which would not have supported significant DNA translocation. Furthermore, it is possible that the mutated subunits formed homohexamers and mixed hexamers more avidly than the wild-type protein on the 16 bp substrate, leading to an over-representation of hexamers containing ≥ 3 mutant subunits for a given input ratio.

In triplex-displacement assays, which provide an independent measure of translocation, hexamers with three or more mutant subunits all failed to show significant displacement activity in assays that required 60 bp of translocation after loading at KOPS, before encountering the triplex (Figure 3C). Hexamers with two WA or two WB mutated subunits showed $\leq 20\%$ of wild-type activity, a marked contrast to their higher activity when activating Xer recombination. ATPase activities measured in the same reactions gave results similar to those in Figure 3B (data not shown). These data suggest a quantitative difference in 'readout' as compared with the recombination assays, perhaps because hexamers with mutated subunits are less able to displace a triplex when they have translocated up to it, as compared with being able to activate recombination after translocation up to the XerCD-*dif* complex. It is unlikely that the difference between assays reflects reduced processivity of translocation with mutated subunits, given that FtsK has to translocate only 60 bp in the triplex assay, whereas most random loadings of FtsK onto DNA in the recombination assay will leave >60 bp of translocation needed to encounter XerCD-*dif*. To test this directly, we

assayed translocation by wild-type and mutated trimers in a single-molecule assay.

Translocases with two catalytically inactive subunits show wild-type translocation velocity

FtsK and its ortholog, SpoIIIE, have been shown to translocate along DNA at ~ 5 kb/s and against forces of up to 60 pN, using single-molecule studies with both optical traps and magnetic tweezers (Saleh *et al*, 2004; Pease *et al*, 2005; Ptacin *et al*, 2008).

We used magnetic tweezers with a DNA molecule (17 kb, 5.7 μ m) tethered between the surface and a magnetic bead. The DNA was maintained in an extended conformation by magnets applying a constant force (Strick *et al*, 1996), to study translocation by hexamers derived from wild-type and mutated trimers. The FtsK variant and ATP were then added and the position of the bead was observed. Looping during translocation, probably arising when the translocase attaches to either the bead or the surface, causes the bead to move towards the surface during translocation. Either release of the loop or a reverse in translocation direction causes the bead to move away from the surface (Figure 4). These events can be distinguished by the rate of reversal of the bead. Between 360 and 930 events per protein were recorded, and the distributions of burst speeds were determined. Translocases derived from the wild-type trimer were first compared with FtsK_{50C}. Both proteins showed the same translocation velocity (data not shown), with translocation events of several microns being common.

Translocases derived from trimers containing a mutated (WA or WB) central subunit were analysed in the same way and the average burst speed was deduced. Surprisingly, they translocated along DNA at rates similar to the wild-type covalent trimer against forces of 1–35 pN. The data shown were obtained at 5 pN and reveal a translocation velocity of ~ 3.3 kb/s. We are confident that these events arise from the translocation of hexamers formed by dimerization of the mutated trimers, because our EM and gel-shift analysis provided no evidence for significant numbers of higher form multimers that could assemble six wild-type subunits into a hexamer (Figures 1 and 2). Also note that monomers and covalent dimers were not active at this concentration. Furthermore, the observation of substantial translocation activity by the covalent trimers with a single-mutated subunit in the XerCD-*dif* recombination reaction, where specific activity can be estimated, reinforces our conclusion that hexamers with mutated subunits in the 2 and 5 positions exhibit wild-type translocation rates. This led us to wonder whether the low activity in the triplex-displacement assays results from a failure to displace the triplex efficiently rather than an impaired ability to translocate up to the triplex. This could mean that the covalent trimer with a central mutant subunit would have an impaired ability to generate the mechanical work required to displace the triplex, despite a normal translocation velocity.

Magnetic tweezers are not a good tool to measure and compare FtsK stall forces as both processivity and events rate dramatically decrease at high force (Saleh *et al*, 2004; Pease *et al*, 2005), while modification of DNA structure at high forces may affect the results. This led us to use another strategy.

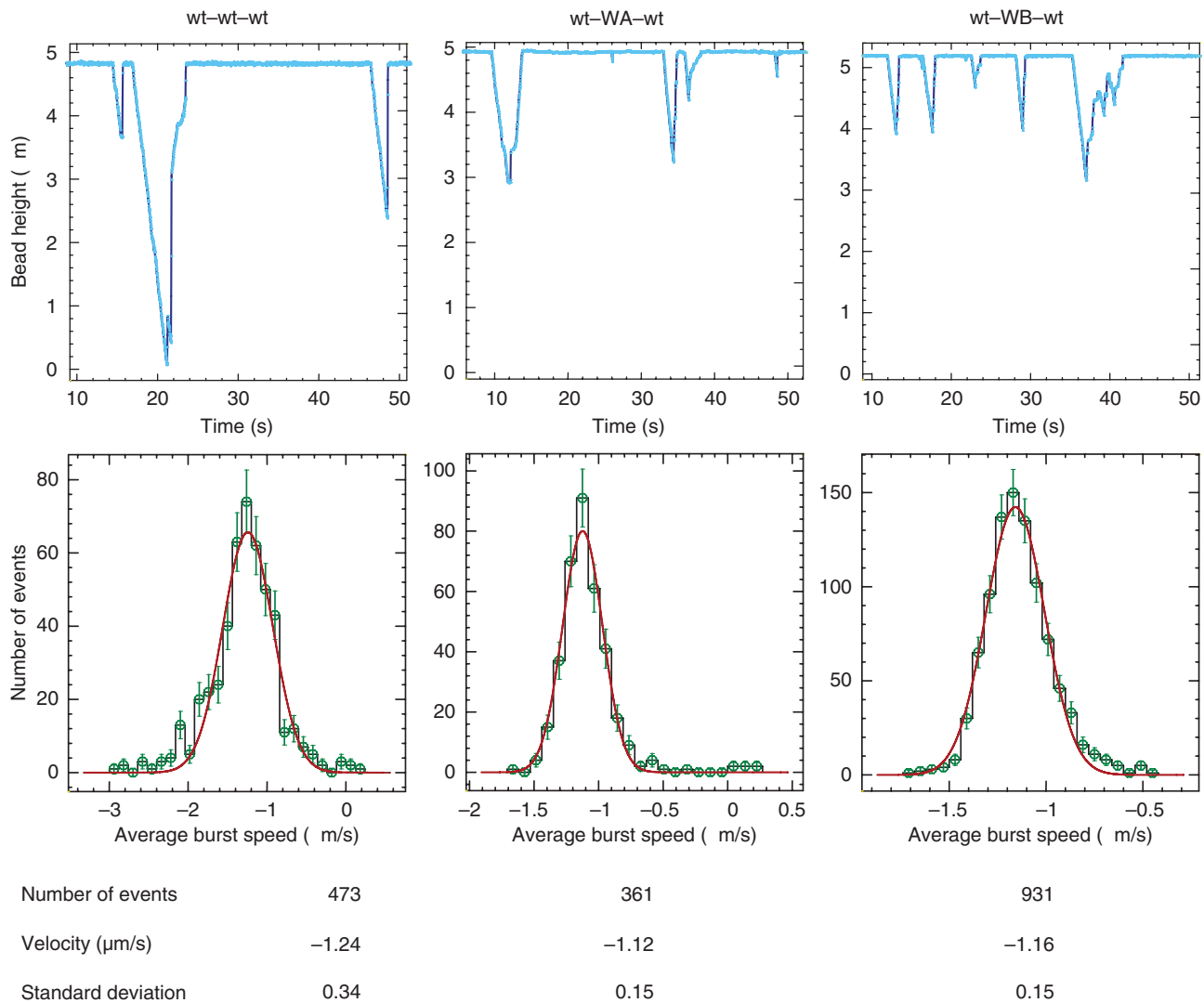


Figure 4 Single-molecule analysis of covalent multimer translocation. Magnetic tweezers were used to study translocation of wild-type and double mutant hexamers, containing WA or WB in the central subunit of the trimers. Upper panels show examples of events obtained for each protein. There is no statistical significance to the relative lengths of the events shown. Light blue, data points; dark blue, fit to the data points. Lower panels show the distribution of average burst speed. The number of events, the average burst speed and its distribution are plotted under each panel. Green, data points, with s.d. (bars); red, Gaussian fits to the data. Note that the WA derivative shown here is biotin tagged and will, therefore, associate with the bead. A His-tagged derivative showed similar velocities. The data shown were obtained at a force of 5 pN and 2 mM ATP.

Trimers with a single catalytic mutation give translocases impaired in displacing a DNA roadblock

To calibrate the relative ability of molecular motors to strip proteins from DNA, we have generated a panel of mutated streptavidins with a wide range of off-rates from biotin (Chivers *et al*, 2010). The streptavidin-biotin interaction is one of the strongest non-covalent interactions characterized, with a K_D of $\sim 4 \times 10^{-14}$ M (Green, 1990). Here, we focused on one stronger streptavidin mutant and one weaker mutant of streptavidin. The stronger streptavidin mutant had a 10-fold lower off-rate from biotin ($<2\%$ spontaneous dissociation from biotin-4-fluorescein in 12 h) than wild-type streptavidin (12% spontaneous dissociation in 12 h). The weaker mutant had accelerated dissociation from biotin conjugates (Supplementary Figure S1), but binding was still strong enough to survive gel electrophoresis.

Therefore, to examine the ability of FtsK to do mechanical work to displace roadblocks on DNA during translocation, we

compared the ability of wild-type and mutated trimers to displace wild-type and mutated streptavidins from biotinylated DNA. A 597 bp DNA fragment biotinylated at one end and with two KOPS sites positioned 230 bp upstream was bound by streptavidin and used in FtsK translocation assays. After 2 min reactions with translocase and ATP, agarose gel electrophoresis was used to analyse the fraction of DNA molecules where streptavidin had been displaced. Excess free biotin was used as a 'sink' to prevent rebinding of displaced streptavidin and 0.1% SDS was used to stop translocation, conditions that do not denature streptavidin (Figure 5). Wild-type hexamers displaced $>70\%$ of streptavidin in 2 min, whereas translocases with two WA or WB mutations displaced $<45\%$ of streptavidin. In all, 30% of the tight-binding streptavidin was displaced in 2 min by the wild-type trimer, whereas $\leq 10\%$ was displaced by the singly mutated trimers. In contrast, the weaker streptavidin was displaced to comparable levels ($\geq 82\%$) by all three proteins,

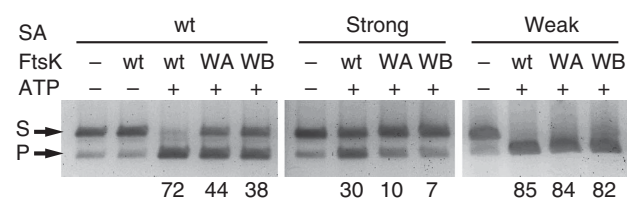


Figure 5 Displacement of the streptavidin 'roadblock' during translocation. Streptavidin displacement activity was followed on a 597 bp biotinylated DNA. Excess streptavidin (wt, strong, or weak binding derivatives) was bound for 30 min and then excess free biotin was added. The indicated trimers (WA and WB contained mutations in the central subunit; 250 nM hexamer) were added, followed by ATP, and the reactions were stopped after 2 min. % displacement (determined by gel electrophoresis; 1.5% agarose/TBE; background subtracted) is indicated. Independent repetitions (≥ 2) gave comparable results. (S) substrate; (P) product.

showing that they have comparable abilities to translocate up to, and displace, a weak roadblock. Taken together, these data show that trimers with one catalytically inactive subunit form translocases as rapid as those formed by wild-type trimers, but which are impaired in displacing strong roadblocks that include triplex DNA and streptavidin bound to biotin.

Discussion

The construction of functional FtsK covalent multimers has provided robust new dsDNA translocases that still recognize KOPS-loading sites on DNA and activate XerCD recombination at *dif*. The use of covalent multimers has revealed new insight into the catalytic mechanism of FtsK and has demonstrated that translocation velocity can be uncoupled from the ability to actively remove roadblocks on DNA. Covalent trimers dimerize efficiently to form biologically active hexamers. We found little formation of higher-order multimers that would complicate the interpretation of the data. Covalent trimers containing catalytic mutations in the central position formed translocases with normal translocation velocities at forces at least up to 10 pN, and were able to translocate over 1 or 2 μ m in a single processive event. Nevertheless, these mutated translocases were compromised in their ability to displace streptavidin from biotinylated DNA or a triplex oligonucleotide. The ability to separate translocation velocity from force generation in displacing a potential 'roadblock' has mechanistic implications and substantiates our belief that we are not forming a wild-type hexamer within a higher-order structure that has catalytically inactive subunits positioned outside of the functional hexamer (Figure 1B). This is the first study that we are aware of that uses covalently linked multimers to study a hexameric DNA translocase, although mechanistic and structural studies of the hexameric ClpX protein-unfolding machine and of RecA have used covalently linked subunits (Martin *et al*, 2005; Chen *et al*, 2008; Glynn *et al*, 2009).

The FtsK translocation mechanism

Historically, three types of mechanism have been considered for the translocation of ring-shaped translocases and helicases (Singleton *et al*, 2007; Thomsen and Berger, 2009). Stochastic mechanisms require that subunits in a ring 'fire' randomly, and such a mechanism has been proposed to operate during protein unfolding by ClpX (Martin *et al*,

2005). A concerted mechanism, in its most rigorous form, requires that all potentially active subunits fire simultaneously. Although a concerted mechanism has been proposed for the translocase action of SV40 T antigen (Gai *et al*, 2004), there is no compelling evidence to believe that such a mechanism operates widely. In contrast, sequential, 'binding change' models for translocation receive extensive support. In these, an asymmetry in the ring, as a consequence of different nucleotide-binding states, is propagated around the ring as DNA or RNA is translocated through the ring. The F1 ATPase provides the paradigm for such a rotary motor, although in this case only three of the six subunits in the ring are catalytically active (Zhou and Boyer, 1993); experimental data derived from studies of a number of helicases and translocases lend support to such a mechanism (reviewed in Singleton *et al*, 2007; Enemark and Joshua-Tor, 2008). Structures of the T7 gp4 helicase in the absence of DNA revealed asymmetric nucleotide-binding states within the ring and led to a 'hand-off' model in which each subunit changes its nucleotide contacts with each catalytic step (Singleton *et al*, 2000; Figure 6A). Subsequent studies of the same protein led to the conclusion that the catalytic mechanism required that all six subunits be catalytically active for translocation to occur, as doping of wild-type protein with a catalytic mutant had an immediate and dramatic effect on DNA-dependent nucleotide hydrolysis; nevertheless, unscheduled ATP hydrolysis in the absence of DNA continued with stochastic firing of catalytically competent subunits (Crampton *et al*, 2006).

Subsequently, a rotary hand-off translocation mechanism for FtsK was proposed on the basis of structural studies, with the suggestion that 2 bp are translocated per catalytic step, with a rotational resetting after each step to accommodate the difference between the normal 10.5 bp per helical turn of B DNA and the 12 bp proposed to be translocated in a complete hexamer catalytic cycle (Massey *et al*, 2006). This resetting can explain the modest level of (+) DNA supercoiling that accumulates ahead of the translocase (Saleh *et al*, 2005). If this mechanism operates during FtsK translocation, the observation that hexamers with two catalytic mutations show normal velocity suggests that the mechanism can accommodate a now substantial rotational (or translational) resetting to locate the next active subunit. Such a major resetting without influence on translocation velocity seems unlikely and is inconsistent with a strict hand-off sequential mechanism. It would also imply that nucleotide binding or release rather than the chemical steps would be rate limiting in such semi-sequential translocation.

More recently, structural studies of E1 papillomavirus helicase (Enemark and Joshua-Tor, 2006) and RNA translocase Rho (Thomsen and Berger, 2009) revealed more incisive mechanistic insight into the translocation process, because the structures showed the asymmetry and different nucleotide-binding states required for a sequential-binding site change mechanism. In each of these cases, a spiral staircase of loops protrudes into the central channel, with each loop grasping the phosphodiester backbone, in a way that can help propel the nucleic acid with respect to the protein. Both of these structures support an 'escorted translocation' mechanism, rather than a hand-off mechanism (Figure 6). In the Rho mechanism, six nucleotides of ssRNA occupy the channel, with each nucleotide potentially being bound by a loop

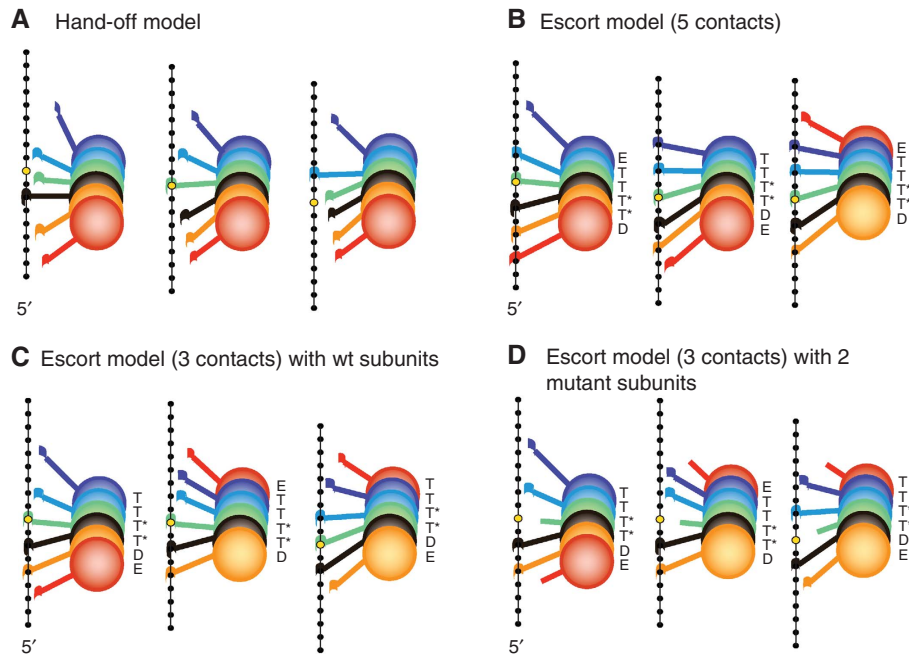


Figure 6 Translocation models. Scheme of FtsK translocation on DNA. The six subunits are drawn as spheres with a paddle representing the loop(s) in the α and/or β domains interacting with DNA nucleotides (black dots). One nucleotide is taken as a reference to show relative movement (yellow dot). Subunits inactivated by WA or WB catalytic mutations are indicated by a shorter paddle without the hook. Note that to simplify the drawing, subunits are represented linearly, but in reality would form a ring around DNA. (A) A sequential hand-off mechanism as applied to FtsK. In this model, all protein subunits change their local association with a given DNA segment, whether it be through specific contacts or not, during each chemical step (Massey *et al*, 2006; Enemark and Joshua-Tor, 2008). Each catalytic step leads to 2 bp dsDNA translocated. (B) Sequential escort model, as applied to FtsK. In this class of model, only one of the subunits changes its DNA contacts at each catalytic step and flexibility in the subunits and their loops allows movement of one subunit from the bottom to the top of the staircase during each catalytic step (Enemark and Joshua-Tor, 2008; Thomsen and Berger, 2009). In this panel, five of the six subunits are contacting every other nucleotide to avoid the substantial overwinding required to contact adjacent nucleotides. This model would require >30 Å movement of a subunit, as it moves from the bottom to the top of the staircase. The indicated nucleotide states of the subunits are; T, ATP bound; T*, ATP transition state; D, ADP bound; E, 'exchange' (nucleotide free) (Thomsen and Berger, 2009). (C) Escort model with three subunit–nucleotide interactions, each occurring every other nucleotide. In this case, the movement of subunits and their loops during transition from the bottom to the top of the staircase would be similar to that for Rho/E1 (~ 15 Å). (D) Escort model in a hexamer containing two mutated subunits. At least two protein–DNA contacts are still feasible in this situation, allowing velocity to remain the same, but reducing the work available to displace roadblocks.

derived from each subunit of the hexamer. After ATP hydrolysis, phosphate is eventually released to give a subunit that loses its strong association with nucleic acid. On ADP release or ATP binding, a subunit loop switches position from the bottom to the top of the staircase, grabbing the 3' nt immediately ahead of the protein and only releasing the nucleotide six catalytic steps (i.e. one complete catalytic cycle) later; this requires a ~ 15 Å movement of the loop. In the Rho structure, which has the nucleotide mimic ADP–BeF₃ bound to each subunit, there are two 'ATP-bound' states (T), two 'ATP-bound' transition states (T*), an 'ADP state' (D) and an exchange (E; nucleotide-free) state, organized in an anti-clockwise manner as one views from the 5' end of the RNA; intersubunit interactions are strongest in the transition state interface and weakest surrounding the exchanging subunit (Thomsen and Berger, 2009). Such a scheme can be applied to FtsK, although the substantial overwinding of duplex DNA in the FtsK hexameric channel, required if there were to be one set of nucleotide contacts per FtsK subunit, seems unlikely. An alternative would be to have one FtsK subunit per two nucleotides, with five contacts at any one time, giving 2 bp translocated per catalytic step and close to normal (or somewhat underwound) B DNA in the channel (Figure 6B). If this were the case, movements in the loop-contacting DNA

in the protein subunit that switches position in a given step would need to span >30 Å of DNA, twice as large as in Rho or E1. However, if only a subset of subunits (Figure 6C) make a DNA contact at any time, then the catalytic cycle could be accommodated with movements similar to those observed for Rho/E1.

An escort model for FtsK translocation can accommodate our observation of normal velocity yet reduced ability to displace roadblocks when two catalytically inactive subunits are introduced into a hexamer. This is because coordinated cooperative interactions within the hexamer and with DNA potentially allow a subunit interface to bypass the active configuration without hydrolysing ATP (Figure 6D). Normal velocity is possible because the chemical steps are not rate limiting and/or because a 'Brownian ratchet' allows the bypass without a diminution in velocity. Nevertheless, a hexamer with two ATP non-hydrolysing subunits would be expected to be able to do less work in displacing a roadblock than a hexamer in which all subunits are catalytically competent, consistent with our observations. It is noteworthy that several hexameric rotary motors have evolved to have a mixture of inactive and active subunits, notably F1 ATPase and dynein. We note that the phage $\phi 12$ P4 hexameric pump, which belongs to the RecA-family, also

seems to use an escort mechanism for ssRNA translocation into the capsid, although in this case it appears that three consecutive nucleotide-bound subunits contact RNA and cooperate in catalysis (Mancini *et al*, 2004; Kainov *et al*, 2008).

Further understanding of the FtsK mechanism will require the combination of structural, biochemical, genetic and single-molecule studies. As yet, we have failed to obtain high-resolution structures of FtsK with DNA, which may reveal the predicted asymmetry in structure and nucleotide-binding state for a sequential-binding change model. Although single-molecule studies have been invaluable in analysing translocation speeds, processivities and directionality of translocation, the rapid speed of translocation has precluded obtaining data on individual catalytic step sizes. Similarly, our attempts to measure the amount of ATP consumed per bp of DNA translocated have had insufficient resolving power, given the speed of translocation (Graham JE and Sherratt DJ, unpublished data).

Finally, the types of study reported here need to be extrapolated to the *in vivo* temporal and spatial behaviour of FtsK in the final stages of chromosome segregation. *In vivo*, FtsK translocation must not be hindered by potential roadblocks on DNA, as FtsK needs to translocate from wherever it loads to XerCD-bound *dif* sites, where translocation stops and the final chromosome unlinking is activated. The ability of FtsK to efficiently displace streptavidin from biotin-DNA suggests that it is an efficient remover of roadblocks *in vivo*. The observation that FtsK hexamers with two mutated subunits can translocate with normal velocity yet are compromised in displacing roadblocks on DNA, also raises the possibility that *in vivo* FtsK translocation can occur in a mode that allows fast translocation without all subunits firing in each complete catalytic cycle, with the full catalytic potential only realized when roadblocks need to be removed during translocation.

Materials and methods

Strains and plasmids

E. coli DS9041 (FtsK_C) was used to express alleles of *ftsK* in *in vivo* recombination experiments (Recchia *et al*, 1999). *E. coli* BL21 was used to overexpress FtsK derivatives. All *ftsK* alleles were cloned into plasmid pBAD24, under an arabinose-inducible promoter and with an N-terminal 6His-tag (Guzman *et al*, 1995). Plasmid pFX142 is a pSC101 derivative containing two *dif* sites (Aussel *et al*, 2002). Kanamycin (50 µg/ml), ampicillin (100 µg/ml) and spectinomycin (25 µg/ml) were used when required. The C-terminus motor of FtsK was cloned from nucleotide 2518 to 3987, or 3990 when the stop codon was required. The monomer was cloned into pBAD24 between *EcoRI* and *HindIII* sites. The dimer contained the first subunit cloned between *EcoRI* and *XbaI* restriction sites, and the second between *XbaI* and *HindIII* sites. The two repeats were joined together by the sequence coding for the 14 amino-acid linker (GGGSEGGGSEGGSG): 5'-GCGGCGGTTCCGAGGGCGGTGGTTTCAG AAGGCGGTTCCGGT-3', followed by the *XbaI* restriction site. Trimers were designed the same way with three fragments cloned between *EcoRI*, *XbaI*, *KpnI* and *HindIII* sites, the last one containing the stop codon before the *HindIII* site.

Toxicity and *in vivo* recombination

Plasmids containing *ftsK* alleles were transformed into DS9041 and cells were grown in LB-glucose to avoid expression from *p_{ara}*. Overnight cultures were diluted 1/100 in LB, and FtsK expression was induced at A₆₀₀ 0.5 by 0.2% arabinose (3 h; 37°C).

Recombination was assayed as in Sivanathan *et al* (2006).

Protein purification

FtsK derivatives were purified as in Massey *et al* (2006), with purification through Talon resin, heparin and Q-HP columns (GE Healthcare). XerC and XerD were purified as in Subramanya *et al* (1997). Streptavidin variants were purified as in Howarth and Ting, 2008.

ATPase, *in vitro* recombination and triplex-displacement assays

These were as described in Massey *et al* (2006) and Sivanathan *et al* (2006), respectively. Reaction buffer (RB) was 25 mM Tris pH 7.5, 10 mM MgCl₂; ATPase assays were carried out on a supercoiled, 9 kb plasmid and *in vitro* recombination on a dimeric plasmid derived from pMin33, which contains a *dif* site.

Electron microscopy and image processing

Protein samples and DNA were mixed at a ratio of 1 hexamer per 66 bp, or 1 per 54 bp for the short DNA and plasmid, respectively, in binding buffer (BB: 25 mM Tris pH 7.5, 30 mM NaCl and 5 mM MgCl₂). The protein-DNA samples were then applied to electron microscope hydrophilic grids coated with carbon film and stained with 2% uranyl acetate. The preparations were examined using a Philips CM120 electron microscope equipped with a LaB₆ filament, with an acceleration voltage of 120 kV. Electron micrographs were taken at a magnification of $\times 45\,000$ at low dose. Selected images were digitized with a step size of 12.5 mm on a Super Coolscan 9000 Nikon. The WEB and SPIDER software package (Frank *et al*, 1996) was used for all image processing. In all, 162 particles on plasmid and 2600 on short DNA were windowed, centred and aligned using the reference-free alignment procedure. For the gold labelling, 3 µl FtsK-DNA complex were applied on glow discharged carbon-coated grids. The grid was washed after 1 min with BB and incubated with Ni-NTA Nanogold (Nanoprobes) solution (5 µM Ni-NTA Nanogold in BB) for 4 min at 25°C. The grid was rinsed twice on drops of BB and stained with 2% uranyl acetate. In all, 1702 particles were windowed and subjected to reference-free alignment and sorted into 10, 50 or 100 classes using K-means clustering (Frank, 1990). A control was made using the same protocol as above but BB was used instead during the labelling step, and no significant difference was observed as for the shape and size of the particles.

Magnetic tweezer assays

Single-molecule experiments were performed with magnetic tweezers (Strick *et al*, 1996). Permanent magnets provided a constant force that varied from 1 to 10 pN in most experiments, although a few experiments were carried out at 35 pN. DNA (pFX355, 17 kb, gift from FX Barre) and surfaces were prepared as described in Lionnet *et al* (2008). Experiments were performed in a buffer containing 2 mM ATP, 10 mM Tris pH 7.6, 150 mM NaCl and 10 mM MgCl₂. FtsK was added until translocation events were observed (typically 50 nM). For each translocation event, speed was obtained from a linear fit to the DNA extension versus time data. For each DNA molecule on a given condition, the individual translocation events were pooled and a histogram of the translocation speed was built, then fitted to a Gaussian function.

Streptavidin-displacement assays

DNA fragments (597 bp) were generated by PCR, using a 5' biotinylated oligonucleotide primer on a plasmid template containing two non-overlapping KOPS pointing towards the biotinylated end. This was first incubated with 0.5 µM streptavidin analogues in RB for 20 min, then 0.1 mM biotin was added as a sink before FtsK (250 nM hexamer). After 2 min incubation, 2 mM ATP was added. Reactions were run at 25°C for 2 min, stopped with 0.1% SDS, 20 mM EDTA pH 8 and loaded on a 1.5% agarose gel in TBE 1 ×. Gels were run at 4°C, 100 V for 90 min and visualized and quantitated as above.

Supplementary data

Supplementary data are available at *The EMBO Journal* Online (<http://www.embojournal.org>).

Acknowledgements

The work in Oxford was supported by the Wellcome Trust (EC, IG, CV-B, MH and DJS) and the BBSRC (CEC). In Paris, the work was

supported by the Wellcome Trust (EC), ANR, IUF and an EU grant, Bionanoswitch (AM and J-FA). We thank many of our mechanistic biochemistry colleagues for invaluable discussions.

References

- Aussel L, Barre FX, Aroyo M, Stasiak A, Stasiak AZ, Sherratt D (2002) FtsK is a DNA motor protein that activates chromosome dimer resolution by switching the catalytic state of the XerC and XerD recombinases. *Cell* **108**: 195–205
- Begg KJ, Dewar SJ, Donachie WD (1995) A new *Escherichia coli* cell division gene, *ftsK*. *J Bacteriol* **177**: 6211–6222
- Bigot S, Saleh O, Cornet F, Allemand J, Barre F (2006) Oriented loading of FtsK on KOPS. *Nat Struct Mol Biol* **13**: 1026–1028
- Chen Z, Yang H, Pavletich NP (2008) Mechanism of homologous recombination from the RecA-ssDNA/dsDNA structures. *Nature* **453**: 489–494
- Chivers CE, Crozat E, Chu C, Moy VT, Sherratt DJ, Howarth M (2010) A streptavidin variant with slower biotin dissociation and increased mechanostability. *Nat Methods* (in press; doi:10.1038/nmeth.1450)
- Crampton D, Mukherjee S, Richardson C (2006) DNA-induced switch from independent to sequential dTTP hydrolysis in the bacteriophage T7 DNA helicase. *Mol Cell* **21**: 165–174
- Dorazi R, Dewar SJ (2000) Membrane topology of the N-terminus of the *Escherichia coli* FtsK division protein. *FEBS Lett* **478**: 13–18
- Enemark EJ, Joshua-Tor L (2006) Mechanism of DNA translocation in a replicative hexameric helicase. *Nature* **442**: 270–275
- Enemark EJ, Joshua-Tor L (2008) On helicases and other motor proteins. *Curr Opin Struct Biol* **18**: 243–257
- Frank J (1990) Classification of macromolecular assemblies studied as ‘single particles’. *Q Rev Biophys* **23**: 281–329
- Frank J, Radermacher M, Penczek P, Zhu J, Li Y, Ladjadj M, Leith A (1996) SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J Struct Biol* **116**: 190–199
- Gai D, Zhao R, Li D, Finkielstein CV, Chen XS (2004) Mechanisms of conformational change for a replicative hexameric helicase of SV40 large tumor antigen. *Cell* **119**: 47–60
- Glynn SE, Martin A, Nager AR, Baker TA, Sauer RT (2009) Structures of asymmetric ClpX hexamers reveal nucleotide-dependent motions in a AAA+ protein-unfolding machine. *Cell* **139**: 744–756
- Graham JE, Sivanathan V, Sherratt DJ, Arciszewska LK (2010) FtsK translocation on DNA stops at XerCD-*dif*. *Nucleic Acids Res* **38**: 72–81
- Grainge I, Bregu M, Vazquez M, Sivanathan V, Ip SC, Sherratt DJ (2007) Unlinking chromosome catenanes *in vivo* by site-specific recombination. *EMBO J* **26**: 4228–4238
- Green NM (1990) Avidin and streptavidin. *Methods Enzymol* **184**: 51–67
- Guzman LM, Belin D, Carson MJ, Beckwith J (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose *pBAD* promoter. *J Bact* **177**: 4121–4130
- Howarth M, Ting AY (2008) Imaging proteins in live mammalian cells with biotin ligase and monovalent streptavidin. *Nat Protoc* **3**: 534–545
- Ip SC, Bregu M, Barre FX, Sherratt DJ (2003) Decatenation of DNA circles by FtsK-dependent Xer site-specific recombination. *EMBO J* **22**: 6399–6407
- Kainov DE, Mancini EJ, Telenius J, Lissal J, Grimes JM, Bamford DH, Stuart DI, Tuma R (2008) Structural basis of mechanochemical coupling in a hexameric molecular motor. *J Biol Chem* **283**: 3607–3617
- Lionnet T, Allemand JF, Revyakin A, Strick TR, Saleh OA, Bensimon D, Croquette V (2008) Single-molecule studies using magnetic traps. In *Single-Molecule Techniques: a Laboratory Manual*, Selvin PR, Ha T (eds). Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press
- Mancini EJ, Kainov DE, Grimes JM, Tuma R, Bamford DH, Stuart DI (2004) Atomic snapshots of an RNA packaging motor reveal conformational changes linking ATP hydrolysis to RNA translocation. *Cell* **118**: 743–755
- Martin A, Baker T, Sauer R (2005) Rebuilt AAA+ motors reveal operating principles for ATP-fuelled machines. *Nature* **437**: 1115–1120
- Massey T, Mercogliano C, Yates J, Sherratt D, Lowe J (2006) Double-stranded DNA translocation: structure and mechanism of hexameric FtsK. *Mol Cell* **23**: 457–469
- Moreau M, McGeoch A, Lowe A, Itzhaki L, Bell S (2007) ATPase site architecture and helicase mechanism of an archaeal MCM. *Mol Cell* **28**: 304–314
- Pease PJ, Levy O, Cost GJ, Gore J, Ptacin JL, Sherratt D, Bustamante C, Cozzarelli NR (2005) Sequence-directed DNA translocation by purified FtsK. *Science* **307**: 586–590
- Ptacin J, Nollmann M, Becker E, Cozzarelli NR, Pogliano K, Bustamante C (2008) Sequence-directed DNA export guides chromosome translocation during sporulation in *Bacillus subtilis*. *Nat Struct Mol Biol* **15**: 485–493
- Recchia GD, Aroyo M, Wolf D, Blakely G, Sherratt DJ (1999) FtsK-dependent and -independent pathways of Xer site-specific recombination. *EMBO J* **18**: 5724–5734
- Reck-Peterson SL, Vale RD (2004) Molecular dissection of the roles of nucleotide binding and hydrolysis in dynein’s AAA domains in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA* **101**: 1491–1495
- Saleh O, Bigot S, Barre F, Allemand J (2005) Analysis of DNA supercoil induction by FtsK indicates translocation without groove-tracking. *Nat Struct Mol Biol* **12**: 436–440
- Saleh O, Pérals C, Barre F, Allemand J (2004) Fast, DNA-sequence independent translocation by FtsK in a single-molecule experiment. *EMBO J* **23**: 2430–2439
- Singleton MR, Dillingham MS, Wigley DB (2007) Structure and mechanism of helicases and nucleic acid translocases. *Annu Rev Biochem* **76**: 23–50
- Singleton MR, Sawaya MR, Ellenberger T, Wigley DB (2000) Crystal structure of T7 gene 4 ring helicase indicates a mechanism for sequential hydrolysis of nucleotides. *Cell* **101**: 589–600
- Sivanathan V, Allen M, De Bekker C, Baker R, Arciszewska L, Freund S, Bycroft M, Löwe J, Sherratt D (2006) The FtsK γ domain directs oriented DNA translocation by interacting with KOPS. *Nat Struct Mol Biol* **13**: 965–972
- Steiner W, Liu G, Donachie WD, Kuempel P (1999) The cytoplasmic domain of FtsK protein is required for resolution of chromosome dimers. *Mol Microbiol* **31**: 579–583
- Strick TR, Allemand JF, Bensimon D, Bensimon A, Croquette V (1996) The elasticity of a single supercoiled DNA molecule. *Science* **271**: 1835–1837
- Subramanya HS, Arciszewska LK, Baker RA, Bird LE, Sherratt DJ, Wigley DB (1997) Crystal structure of the site-specific recombinase, XerD. *EMBO J* **16**: 5178–5187
- Thomsen ND, Berger JM (2009) Running in reverse: the structural basis for translocation polarity in hexameric helicases. *Cell* **139**: 523–534
- Vallee RB, Hook P (2006) Autoinhibitory and other autoregulatory elements within the dynein motor domain. *J Struct Biol* **156**: 175–181
- Watanabe YH, Motohashi K, Yoshida M (2002) Roles of the two ATP binding sites of ClpB from *Thermus thermophilus*. *J Biol Chem* **277**: 5804–5809
- Wiese C, Hinz JM, Tebbs RS, Nham PB, Urbin SS, Collins DW, Thompson LH, Schild D (2006) Disparate requirements for the Walker A and B ATPase motifs of human RAD51D in homologous recombination. *Nucleic Acids Res* **34**: 2833–2843
- Yu XC, Weihe EK, Margolin W (1998) Role of the C terminus of FtsK in *Escherichia coli* chromosome segregation. *J Bacteriol* **180**: 6424–6428
- Zhou JM, Boyer PD (1993) Evidence that energization of the chloroplast ATP synthase favors ATP formation at the tight binding catalytic site and increases the affinity for ADP at another catalytic site. *J Biol Chem* **268**: 1531–1538

Conflict of interest

The authors declare that they have no conflict of interest.



The EMBO Journal is published by Nature Publishing Group on behalf of European Molecular Biology Organization. This article is licensed under a Creative Commons Attribution-NonCommercial-Share Alike 3.0 Licence. [http://creativecommons.org/licenses/by-nc-sa/3.0/]

How the biotin–streptavidin interaction was made even stronger: investigation via crystallography and a chimaeric tetramer

Claire E. CHIVERS, Apurba L. KONER, Edward D. LOWE and Mark HOWARTH¹

Department of Biochemistry, Oxford University, South Parks Road, Oxford OX1 3QU, U.K.

The interaction between SA (streptavidin) and biotin is one of the strongest non-covalent interactions in Nature. SA is a widely used tool and a paradigm for protein–ligand interactions. We previously developed a SA mutant, termed Tr (traptavidin), possessing a 10-fold lower off-rate for biotin, with increased mechanical and thermal stability. In the present study, we determined the crystal structures of apo-Tr and biotin–Tr at 1.5 Å resolution. In apo-SA the loop (L3/4), near biotin's valeryl tail, is typically disordered and open, but closes upon biotin binding. In contrast, L3/4 was shut in both apo-Tr and biotin–Tr. The reduced flexibility of L3/4 and decreased conformational change on biotin binding provide an explanation for Tr's reduced biotin off- and on-rates.

L3/4 includes Ser⁴⁵, which forms a hydrogen bond to biotin consistently in Tr, but erratically in SA. Reduced breakage of the biotin–Ser⁴⁵ hydrogen bond in Tr is likely to inhibit the initiating event in biotin's dissociation pathway. We generated a Tr with a single biotin-binding site rather than four, which showed a similarly low off-rate, demonstrating that Tr's low off-rate was governed by intrasubunit effects. Understanding the structural features of this tenacious interaction may assist the design of even stronger affinity tags and inhibitors.

Key words: avidin, biotin, protein engineering, protein–ligand interaction, streptavidin, traptavidin.

INTRODUCTION

The capture of the small molecule biotin (vitamin H/vitamin B₇) by the bacterial protein SA (streptavidin) is both a powerful tool in biology and a model system for the study of high-affinity protein–ligand interactions. The key features are the femtomolar affinity, high specificity, high on-rate and the resilience of SA to pH, temperature and denaturant [1–3]. It is also important that biotin protein ligases can attach biotin to specific lysine residues *in vitro* or in living cells [4,5]. SA's applications include imaging [6], nano-assembly [7] and pre-targeted cancer immunotherapy [8]. As a favoured model in biophysics and protein engineering, this protein–ligand interaction has been investigated by at least 200 point mutations of SA [3,9–11], more than 30 alternative small molecule ligands [1,12–14], computational simulation [15–18], crystallography [2] and single-molecule force dynamics [19].

Despite the strength of SA–biotin binding, the interaction can and does fail under challenging conditions. Biotin dissociation becomes significant and sometimes rapid in response to the low pH of the endosomes [20], the high temperatures used for DNA amplification [21], attachment to nanoparticles [22], and shear forces from flow [23]. The low off-rate of wild-type SA for biotin conjugates was improved upon recently, when we found that the S52G R53D mutant of SA, named Tr (traptavidin) [23a], had a more than 10-fold slower biotin dissociation than wild-type SA, as well as enhanced mechanical and thermal stability [24]. The stronger binding of Tr enabled us to generate a challenging roadblock for one of the fastest known linear molecular motors, FtsK, in order to evaluate the co-ordination of firing around the hexameric ring as FtsK translocated on DNA [24,25].

To understand the basis of Tr's resilient binding and stability to high temperatures, in the present study we solved the crystal structures of both apo-Tr and biotin–Tr and compared these with existing structures of SA. In addition we purified a monovalent

Tr, to investigate the role of intersubunit effects in Tr stability and also to characterize a tool which enables ultra-stable binding of biotinylated targets without cross-linking.

EXPERIMENTAL

Tr purification

Tr protein (S52G R53D core streptavidin with a C-terminal His₆ tag, GenBank® accession number GU952124) was purified from *Escherichia coli* by inclusion body isolation, refolding and Ni-NTA (Ni²⁺-nitrilotriacetate) chromatography, as described previously [24]. Prior to the crystal tray set-up, we performed size-exclusion chromatography (XK 26 column, GE Healthcare), eluting with 50 mM Tris/HCl (pH 7.5), 0.5 M NaCl and 10 mM EDTA.

Tr1D3 (monovalent Tr) was generated by the mixed refolding of Tr and D (dead streptavidin) subunits (N23A S27D S45A core SA with no His₆ tag). Purification was as previously described for monovalent SA [26], except that the order of SDS/PAGE mobility was reversed (mobility of Tr2D2 > Tr1D3 > D4). D4 (a tetramer of dead streptavidin subunits) was generated from the refolding of D from inclusion bodies into PBS and ammonium sulfate precipitation [26].

Apo-Tr crystallization

Crystals were obtained by the sitting-drop vapour-diffusion method at 291 K. Apo-Tr crystals [space group *I*-4₁; *a* = *b* = 57.59 Å (1 Å = 0.1 nm), *c* = 183.35 Å, with two Tr monomers in the asymmetric unit] were obtained from a 4 µl drop of 19 mg/ml solution and a reservoir solution of 12% (v/v) PEG [poly(ethylene glycol)] 8000, 9% (v/v) ethylene glycol and 0.1 M Hepes (pH 7.5). Crystals appeared after 2 days and reached

Abbreviations used: D, dead streptavidin subunit; DPI, diffraction-data precision indicator; D4, a tetramer of dead streptavidin subunits; PEG, poly(ethylene glycol); rmsd, root mean square deviation; SA, streptavidin; Tr, traptavidin; Tr1D3, monovalent Tr; Tr4, tetravalent Tr.

¹ To whom correspondence should be addressed (email mark.howarth@bioch.ox.ac.uk).

The structural co-ordinates reported will appear in the PDB under accession codes 2Y3E and 2Y3F.

optimum size after 4 days. Prior to data collection, crystals were briefly soaked in a cryoprotectant solution of 12 % (v/v) PEG 8000, 30 % (v/v) ethylene glycol and 0.1 M Hepes (pH 7.5) before immersion into liquid nitrogen.

Biotin–Tr crystallization

Biotin-bound Tr was obtained by incubation of Tr with biotin (Acros Organics), in 4-fold molar excess compared with the number of binding sites, at 4°C overnight. Biotin-bound crystals (space group $P4_22_12_1$; $a = b = 57.34$ Å, $c = 77.55$ Å, with one monomer in the asymmetric unit) were obtained from a 4 µl drop of 19 mg/ml solution and a reservoir solution of 27 % (v/v) PEG 4000, 0.25 M MgCl₂ and 0.1 M Tris/HCl (pH 8.5). Crystals appeared after 2 days and reached optimum size after 4 days. Prior to data collection, crystals were briefly soaked in a cryoprotectant solution of 27 % (v/v) PEG 4000, 0.25 M MgCl₂ and 25 % glycerol and 0.1 M Tris/HCl (pH 8.5) before immersion into liquid nitrogen.

Diffraction data collection

Crystallographic data were collected at 100 K, using an Oxford Cryosystems 700 series Cryostream on an ADSC Quantum 315 CCD (charge-coupled device) detector with an oscillation range of 0.5° at beamline IO2 at the Diamond Light Source, Harwell, U.K.

Structure solution and refinement: apo-Tr

Data were indexed and integrated using MOSFLM, and scaled and merged using SCALA from the CCP4 program suite [27]. The structure was phased by molecular replacement using a wild-type core SA search model (PDB code 1SWB) [28] with the program Phaser. The crystal contained two monomers in the asymmetric unit. A total of 5 % of the reflection data were set aside using the Freerflag program in CCP4 and used for the calculation of R_{free} . The model was built in Coot and refined in PHENIX refine, with incorporation of the twinning operator ($-h, k, -l$) with a twinning fraction of 49.7 %. Throughout the refinement, all results were included from 27.47 Å resolution to the highest limit (1.45 Å) and anisotropic temperature factors were refined. The model was evaluated with MolProbity [28b], which gave Ramachandran statistics with 98 % of residues in favoured regions and no outliers. The DPI (diffraction-data precision indicator) indicated that the agreement between the model and the X-ray data for apo-Tr was 0.056 Å. The refinement statistics are shown in Table 1.

Structure solution and refinement: biotin–Tr

The results were indexed and integrated using MOSFLM, and scaled and merged using SCALA from the CCP4 program suite. Intensities were converted into structure factors using the truncate program in CCP4 and no signs of twinning were observed in the results. The structure was phased by molecular replacement using the apo-Tr search model and Phaser. The crystal contained one monomer in the asymmetric unit. A total of 5 % of the reflection data were set aside using the Freerflag program in CCP4 and used for the calculation of R_{free} . Model building was performed in Coot and refinement, including rounds of simulated annealing refinement, in PHENIX refine. Throughout the refinement, all data were included from 38.78 Å resolution to the highest limit

Table 1 Data collection and refinement statistics for crystal structures of apo-Tr and biotin–Tr

The values in parentheses are for the highest-resolution shell. n/a, not applicable.

	Apo-Tr (PDB code 2Y3E)	Biotin–Tr (PDB code 2Y3F)
Data collection		
Space group	$I 4_1$	$P 4_2 2_1 2$
Cell dimensions		
a, b, c (Å)	57.59, 57.59, 183.35	57.34, 57.34, 77.55
α, β, γ (°)	90.00, 90.00, 90.00	90.00, 90.00, 90.00
Resolution (Å)	27.47–1.45 (1.53–1.45)	38.78–1.49 (1.57–1.49)
R_{merge}	0.052 (0.619)	0.069 (0.085)
$I/\sigma I$	13.6 (2.0)	26.0 (19.6)
Completeness (%)	97.8 (87.1)	100.0 (100.0)
Multiplicity	3.6 (3.2)	7.0 (7.1)
Refinement		
Resolution (Å)	27.47–1.45	32.12–1.49
No. reflections	48884	21581
$R_{\text{work}}/R_{\text{free}}$	0.134/0.178	0.135/0.152
No. atoms		
Protein	1855	954
Biotin	n/a	16
Glycerol	18	18
Water	321	142
B factors (mean of all atoms, Å ²)		
Protein	33.41	10.10
Biotin	n/a	5.83
Glycerol	44.14	28.31
Water	38.19	27.92
Rmsd		
Bond length (Å)	0.005	0.009
Bond angle (°)	0.93	1.48
Twinning		
Twin operator	$-h, k, -l$	n/a
Twin fraction	0.497	n/a

(1.49 Å) and anisotropic temperature factors were refined. The presence of a well defined biotin model in the binding pocket was clearly visible in the initial $F_{\text{obs}} - F_{\text{calc}}$ electron density maps. Biotin co-ordinates from wild-type core SA with biotin at 1.4 Å (PDB code 1MK5) [29] were used as a template for the refinement and the weighting of co-ordinate refinement was optimized by PHENIX refine as the refinement progressed. The model was evaluated with MolProbity [28b], which gave Ramachandran statistics with 98 % of residues in favoured regions and no outliers. The DPI indicated that the agreement between the model and the X-ray data, for biotin–Tr was 0.051 Å. The refinement statistics are shown in Table 1. To assess whether model bias was introduced into the biotin–Tr model by using the apo-Tr structure as a search model for molecular replacement, a simulated annealing composite omit map was calculated using PHENIX AutoBuild [30], with the starting phases provided by the apo-Tr structure.

Structure analysis

Structural overlays and colouring according to B factor were done in PyMOL (DeLano Scientific; <http://www.pymol.org>). Alignments were calculated from main-chain atoms of apo-Tr chain A, biotin–Tr, apo-core-SA at pH 4.5 (PDB code 1SWA and 1SWC), apo-core SA at pH 7.5 (PDB code 1SWB) and biotin–core-SA at pH 4.5 (PDB code 1SWE) [28]. Plots of mean B factors of main-chain atoms against residue number were produced using the BAVERAGE program in CCP4. The DPI was assessed using the program SFCHECK from CCP4.

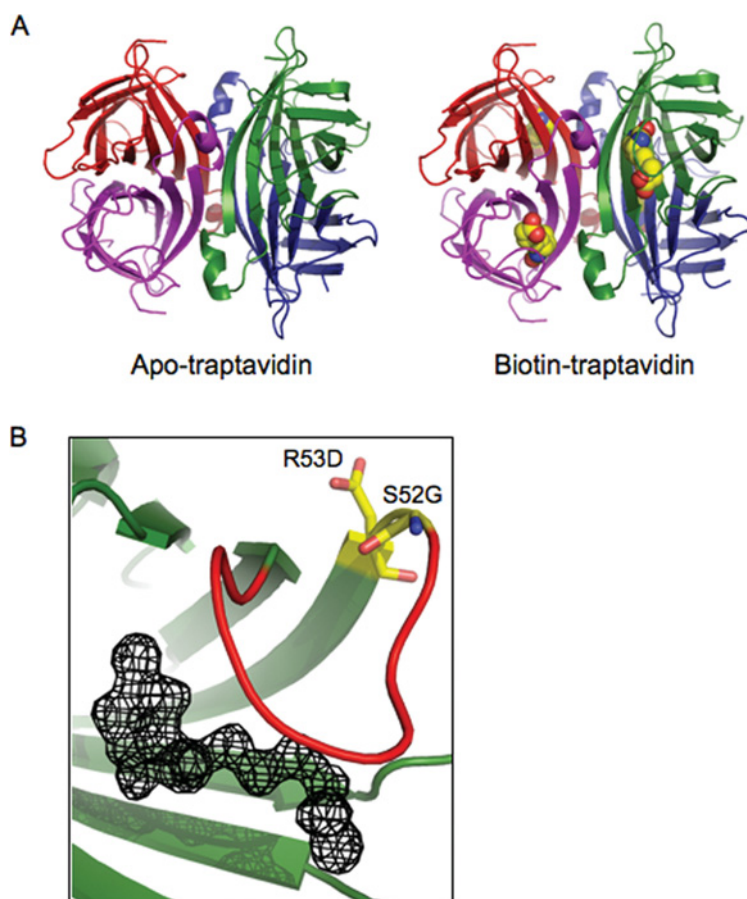


Figure 1 Crystal structures of Tr

(A) Structure of the complete tetramers of apo-Tr (left-hand panel) and biotin-Tr (right-hand panel), with each subunit of the tetramer in a different colour. Biotin bound to each subunit is shown in space-filling mode. (B) $F_{\text{obs}} - F_{\text{calc}}$ map of the biotin-binding pocket in biotin-Tr, showing the electron density for biotin, with the protein overlaid in cartoon format. L3/4 is shown in red and the residues mutated in Tr are in stick format.

Biotin-conjugate off-rate assay

The off-rate of biotin-4-fluorescein was measured from the fluorescence increase on unbinding at 37 °C with competing free biotin, as previously described [26,31].

Thermostability analysis

A 3 μM sample of dead streptavidin, tetravalent Tr or monovalent Tr in PBS was heated at the indicated temperature for 3 min in a DNA Engine[®] Peltier Thermal Cycler (Bio-Rad). SDS/PAGE (8% gel) was performed to discriminate the monomers as described previously [26]. The band intensities were quantified using a ChemiDoc XRS imager and QuantityOne 4.6 software (Bio-Rad). Percentages were calculated as $100 \times (\text{summed intensity of monomer bands at the indicated temperature} - \text{summed intensity of monomer bands at } 25^\circ\text{C}) / (\text{summed intensity of monomer bands after } 95^\circ\text{C with SDS})$.

RESULTS

L3/4 of Tr is closed even without biotin

We determined the crystal structures at high resolution of Tr with and without biotin (1.45 Å for apo-Tr, PDB code 2Y3E; 1.49 Å for biotin-Tr, PDB code 2Y3F) (Table 1). Both structures were solved

by molecular replacement; the apo-SA structure (PDB code 1SWB) was used as a search model for the apo-Tr structure [28]. The subsequent apo-Tr structure was used as the search model for biotin-bound Tr. To determine whether model bias was introduced into the biotin-Tr structure by using the apo-Tr structure as a search model for molecular replacement, a simulated annealing composite omit map was calculated. Alignment of this map with the biotin-Tr and apo-Tr models revealed the final biotin-Tr model fitted the omit map much better (correlation coefficient of 0.73 for side-chain atoms) than the apo-Tr model (correlation coefficient of 0.47 for side-chain atoms) (Supplementary Figure S1 at <http://www.BiochemJ.org/bj/435/bj4350055add.htm>), thus indicating that the solution of the biotin-Tr structure was not biased by the search model.

We observed two subunits in the asymmetric unit for apo-Tr and one subunit in the asymmetric unit for biotin-Tr. We applied symmetry transformations to view the complete tetramer (Figure 1A). SA/Tr subunits are eight-stranded antiparallel β -barrels, with the biotin-binding site located at one end of the barrel. Each subunit binds one biotin molecule and the subunits come together to form a tetramer that can be considered a dimer of dimers. The electron density for the biotin bound to Tr was unambiguous (Figure 1B). A striking difference between Tr and SA was seen in the loop connecting β -strands 3 and 4 (L3/4). L3/4 is commonly disordered in apo-SA ('open' conformation)

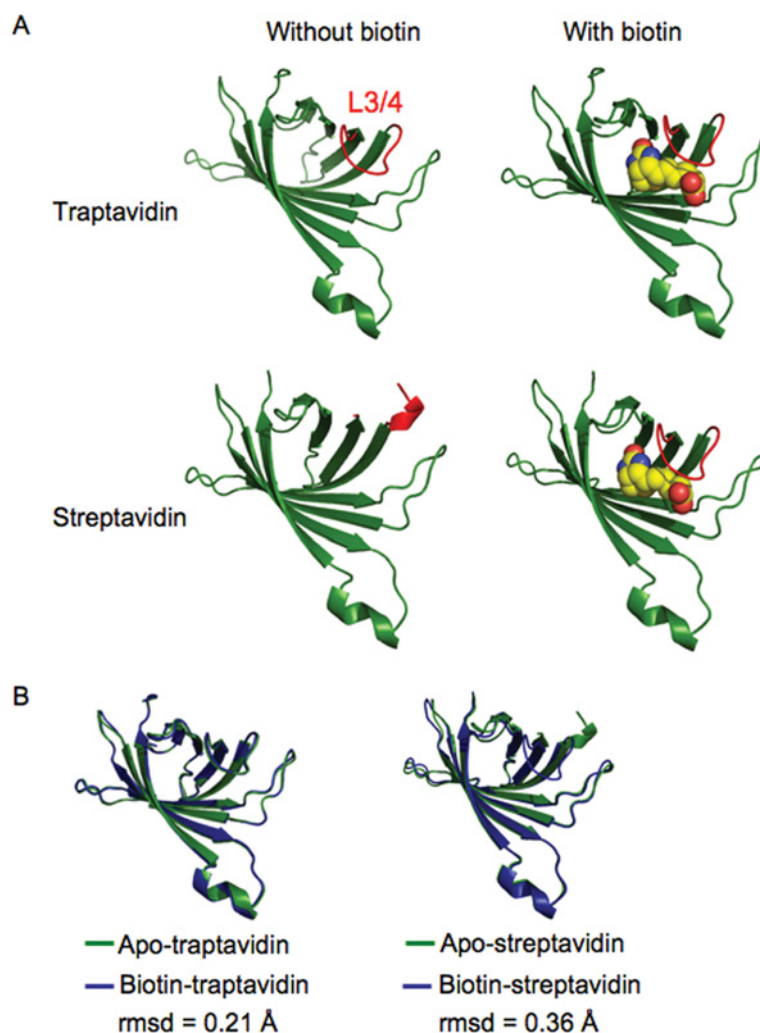


Figure 2 L3/4 of Tr is shut with and without biotin

(A) Comparison between the structure of individual subunits of Tr or SA, with or without biotin. Tr (top row) has a well-defined closed L3/4 (coloured red) in both apo- (left-hand panel) and biotin-bound (right-hand panel) structures. For SA (bottom row) L3/4 is disordered in apo-SA (left-hand panel) (1SWA chain B) but ordered in biotin-SA (1SWE chain D) (right-hand panel) [28]. Biotin is shown in space-filling mode. (B) Conformational changes on biotin binding in Tr (left-hand panel) and SA (right-hand panel), shown as individual subunits. The apo-structure in green is overlaid with the biotin-bound structure in blue.

[28], but becomes ordered on biotin binding, closing over the binding pocket and forming a 'lid' over the bound biotin ('closed' conformation) (Figure 2A). However, in apo-Tr, all L3/4 of the tetramer was already in the 'closed' conformation. L3/4 remained in this closed conformation in biotin-Tr (Figure 2A). Apo-SA structures 1SWA and 1SWB had only one out of four subunits, whereas 1SWC had two out of four subunits with an ordered L3/4 in the tetramer (1SWA, 1SWB and 1SWC have four subunits in the asymmetric unit) [28]. We emphasize that the fact that apo-Tr had two subunits in the asymmetric unit does not mean that the third and fourth subunits were less clearly defined than if there were four subunits in the asymmetric unit; rather the crystallographic symmetry tells us that the third and fourth subunits are identical with the first and second subunits.

Structural alignments of these apo-SA structures with apo-Tr revealed the closed loops in the apo-SA structures are comparable with L3/4 in apo-Tr; the rmsd (root mean square deviation) from apo-Tr chain A for residues 45–52 is 0.24 Å for 1SWA chain A and 0.20 Å for 1SWB chain A. The presence of one 'closed' loop in

the apo-SA tetramer suggests that L3/4 is dynamic in the absence of biotin. L3/4 in the 1SWC apo-SA structure is an exception, with an alternative open conformation, projecting away from the binding site, and is not comparable with L3/4 in apo-Tr (rmsd 2.2 Å compared with 1SWC chain B) [28].

The whole structure of Tr is preformed for biotin binding

SA does not show co-operativity in biotin binding between the four subunits [32], but there is a substantial conformational change when biotin binds, involving the flattening and tighter wrapping of the β -barrels, altered dimer-dimer packing and loop ordering [33]. This conformational change is consistent with an observed increase in thermostability of the biotin-bound tetramer [34]. Upon aligning the apo-Tr and biotin-Tr structures, we found that the conformational change in Tr upon biotin binding is much smaller than for SA (rmsd 0.21 Å for Tr, 0.36 Å for SA) (Figure 2B). Differences in the apo- and biotin-bound conformations of SA were clearly shown by calculating alignments of the loops connecting the β -strands (Table 2). It was

Table 2 Structural alignments of loop regions

(a) Loops of apo-Tr were aligned to biotin-Tr, apo-SA and biotin-SA. Residues comprising each loop are shown in parentheses. (b) Loops of apo-SA were aligned to biotin-SA, apo-Tr and biotin-Tr. Only main-chain atoms were used to calculate the alignments. *The loop of 1SWA was compared only with the resolved residues of 45 and 49–52, since 46–48 were disordered.

(a)

Apo-Tr (chain A) aligned with	Rmsd (Å) of alignment of loop in:		
	Biotin-Tr	Apo-SA (1SWA chain B)	Biotin-SA (1SWE chain D)
L1/2 (23–26)	0.050	0.20	0.21
L2/3 (33–38)	0.14	0.23	0.15
L3/4 (45–52)	0.25	2.3*	0.25
L4/5 (62–70)	0.20	0.20	0.22
L5/6 (79–88)	0.18	0.28	0.18
L6/7 (98–102)	0.14	0.22	0.37
L7/8 (113–121)	0.20	0.23	0.24
Overall structure	0.21	0.29	0.35

(b)

Apo-SA (1SWA chain B) aligned with	Rmsd (Å) of alignment of loop in:		
	Biotin-SA (1SWE, chain D)	Apo-Tr (chain A)	Biotin-Tr
L1/2 (23–26)	0.27	0.20	0.18
L2/3 (33–38)	0.23	0.23	0.22
L3/4 (45–52)	2.6*	2.3*	2.3*
L4/5 (62–70)	0.25	0.20	0.25
L5/6 (79–88)	0.31	0.28	0.27
L6/7 (98–102)	0.42	0.22	0.26
L7/8 (113–121)	0.17	0.23	0.15
Overall structure	0.36	0.29	0.25

not only L3/4 which moved less upon biotin binding in Tr than SA, but also five out of six of the other loops.

Changes in the flexibility of Tr

To analyse the flexibility and dynamics of apo-Tr and biotin-Tr, we plotted the mean *B* factors for main-chain atoms against residue number (Figure 3). Relative *B* factors within a structure are informative, but it is not straightforward to compare directly the absolute values of *B* factors between SA and Tr, because of different resolutions, data collection conditions and refinement methods.

The *B* factors for each of the two monomers in the asymmetric unit of apo-Tr were plotted to indicate the internal variation. This showed the expected low dynamics in the β -sheet regions and some flexibility in all the loop sections (Figure 3A). The main difference between chains A and B of apo-Tr occurs at L5/6 and may result from one chain being involved in crystal contacts at this loop. There is some flexibility in L3/4 for apo-Tr, but clearly it is less than in apo-SA structures where the loop is not resolved. If L3/4 in apo-Tr did not have some flexibility then it is likely that the Tr on-rate would be too low for practical application. The most striking change in the biotin-Tr *B* factors is the loss of flexibility in L3/4, with values comparable with adjacent β -sheets 3 and 4 (Figure 3B), which is consistent with the exceptionally low biotin dissociation rate.

Hydrogen bonding to biotin and Ser⁴⁵ in Tr and SA

The extensive hydrogen-bonding network from SA to biotin plays a key role in generating the large ΔG° of biotin binding [2]. We compared all of the hydrogen bonds to biotin for Tr and SA, paying particular attention to Ser⁴⁵ and Asn⁴⁹, which are present on L3/4 close to the residues mutated in Tr. Factoring in the variation seen in the different chains of the

different structures of SA, the hydrogen-bond lengths to biotin are comparable between Tr and SA (Supplementary Figure S2 at <http://www.BiochemJ.org/bj/435/bj4350055add.htm>).

The published structures of SA allude to the probable dynamic nature of this hydrogen-bond network to biotin. In some crystal structures of SA (1SWE), each Ser⁴⁵ residue has its side-chain hydroxy group directed towards the biotin N3' nitrogen, and hydrogen bonds are formed (although the distance in subunits 1 and 4 in 1SWE is at the limit for a hydrogen bond between such atoms, at 3.2 Å) (Figure 4A and Supplementary Figure S2). However, in the SA structure of 1SWD, the Ser⁴⁵ side chains point away from the N3' nitrogen and so cannot hydrogen bond (Figure 4A and Supplementary Figure S2). In the biotin-Tr structure, the Ser⁴⁵ side chain is clearly orientated towards the N3' nitrogen (Figure 4A), with a hydrogen bond length of 3.0 Å.

Gly⁵² in Tr undergoes little change upon biotin binding (Figure 4B). The peculiar importance of the S52G mutation for biotin-binding stability is likely to relate to Ser⁵² stabilizing an alternative open conformation of L3/4, thereby destabilizing the Ser⁴⁵ hydrogen bond to biotin. In this rival conformation, Ser⁵²'s main-chain carbonyl forms a hydrogen bond to the main-chain N–H of Ser⁴⁵, while the Ser⁵² side chain makes a hydrogen bond to the main-chain N–H of Ala⁴⁶ (Figure 4B). This competing pairing is seen in apo-SA structures with an undefined L3/4 (1SWC) [28], but could well form transiently even in the presence of biotin, so contributing to the rare biotin dissociation events from SA.

The network of polar and non-polar interactions made to biotin by SA changes biotin's structure from the conformation reported to apply in solution [35]. In avidin and SA there are three residues that are positioned to hydrogen bond to the carbonyl oxygen of biotin, suggesting that this oxygen is partially charged and that this polarization may contribute to the exceptional affinity of biotin binding [16]. The bond lengths and bond angles of biotin were

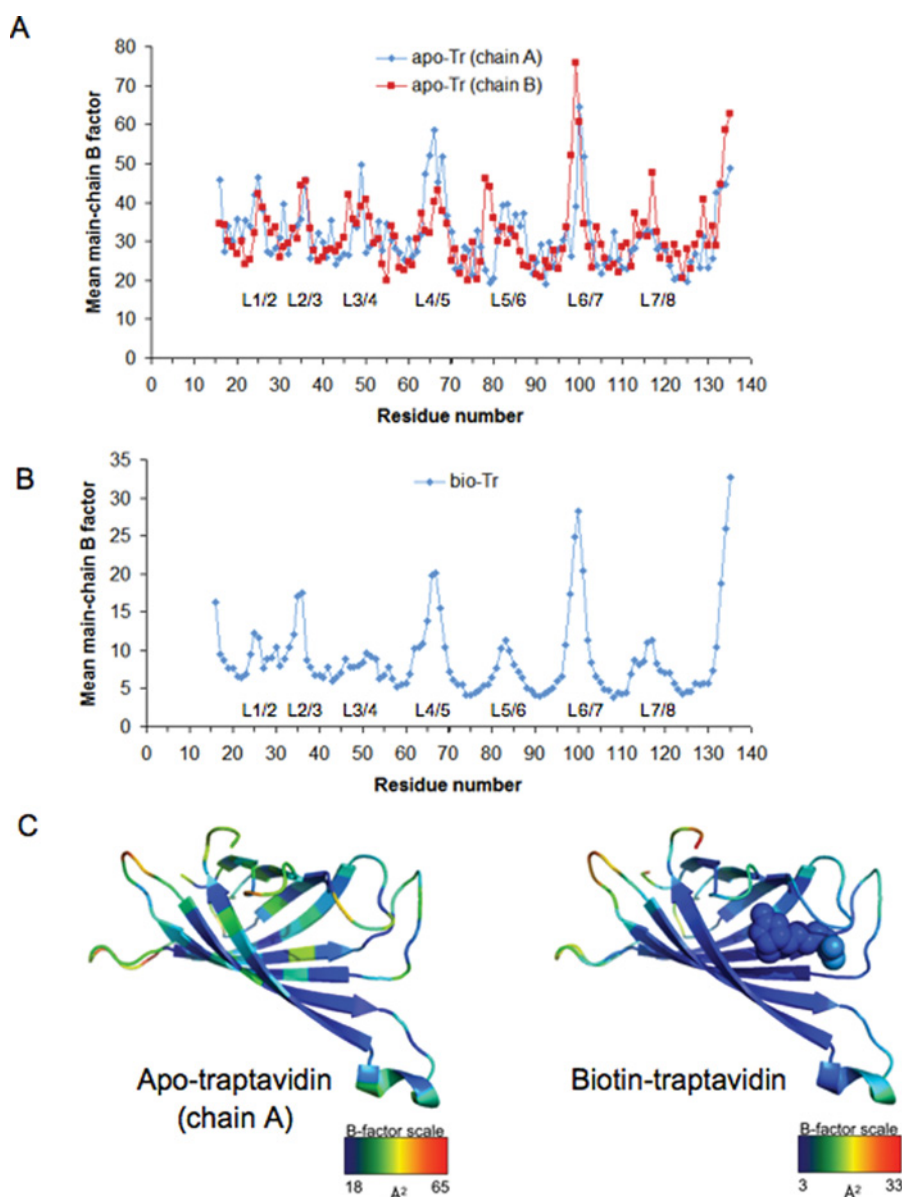


Figure 3 Tr flexibility

(A) B factors (indicating dynamics and flexibility) averaged over main-chain atoms for each amino acid residue are plotted for apo-Tr, with the two chains of the asymmetric unit shown separately. The regions corresponding to loops are labeled L. (B) Mean main-chain B factors for each residue of biotin-Tr, with loops labelled. (C) Heat map showing mean main-chain B factors for an individual subunit of apo-Tr (left-hand panel, chain A) or biotin-Tr (right-hand panel), with biotin shown in space-filling mode. Red indicates the most flexible and blue the least flexible region of the particular subunit.

similar between SA and Tr (results not shown), indicating that changes in biotin's electronic structure in the binding site are not likely to contribute to the difference in biotin off-rate.

Generating monovalent Tr shows that altered intersubunit contacts do not explain Tr's increased stability

We created tetramers with exactly one Tr subunit and three 'dead' subunits which do not bind biotin [26], to give monovalent traptavidin (Tr1D3) (Figure 5A). We previously showed that monovalent SA had equivalent biotin-binding stability to tetravalent SA [26]. Here we used Tr1D3 to explore whether the increased stability of Tr over SA depended upon the presence

of neighbouring Tr subunits. The off-rate of biotin-4-fluorescein was compared for Tr4, Tr1D3 and tetravalent SA (Figure 5B). Tr1D3 had an equivalent off-rate to that of Tr4, showing that it is the binding by the Tr subunit and not any altered intersubunit interaction that is the dominant factor in Tr's improved biotin-binding stability.

Tr stays tetrameric and remains bound to biotin conjugates at higher temperatures than SA [24]. We tested the thermostability of Tr1D3 by incubating it at a range of temperatures and testing by SDS/PAGE what fraction remained tetrameric (Figure 5C). Monovalent SA is known to show thermostability equivalent to that of wild-type SA [26]. We compared Tr1D3 to Tr4 and a tetramer composed entirely of dead subunits (D4). Tr4 was ~10°C more stable than D4. The thermostability of Tr1D3

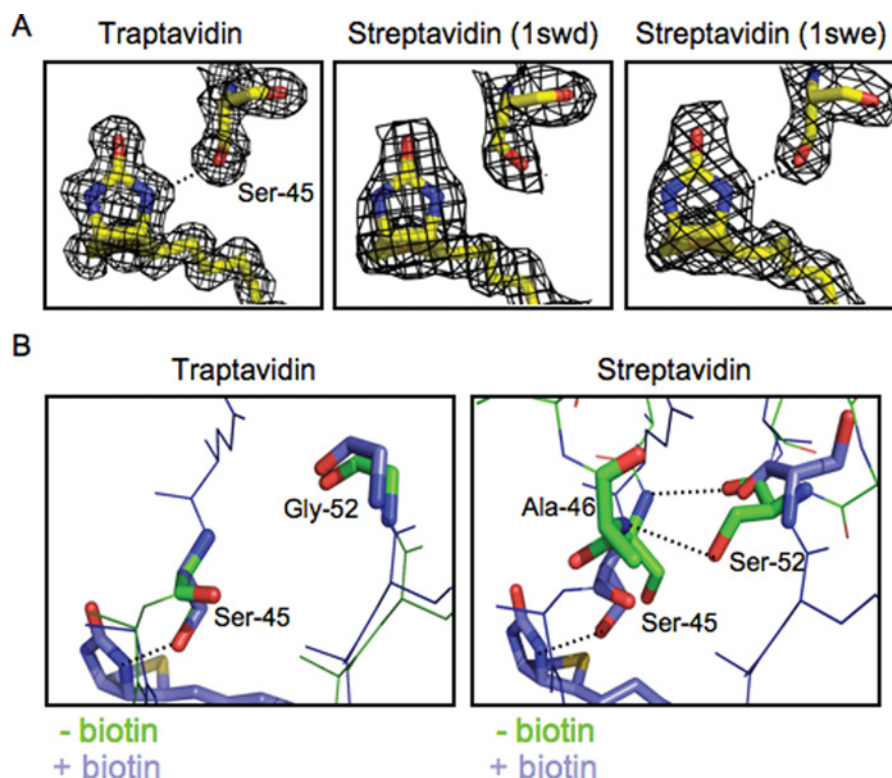


Figure 4 Hydrogen bonding by Ser⁴⁵ in Tr

(A) The electron density for biotin and Ser⁴⁵ surrounds a stick model of the atoms involved, for Tr (left-hand panel) and two different streptavidin structures (1SWD in the middle panel and 1SWE in the right-hand panel). Hydrogen bonds are indicated by a broken line. (B) The rival hydrogen bonding by Ser⁴⁵. An overlay of the L3/4 region, with key residues highlighted in stick format. For Tr (left-hand panel) with and without biotin L3/4 is closed, with no hydrogen bond between Ser⁴⁵ and Gly⁵². For SA (right-hand panel) without biotin and with an undefined L3/4 (1SWC, chain B), Ser⁵² forms hydrogen bonds to Ser⁴⁵ and Ala⁴⁶ (indicated by broken lines). When biotin is bound and the loop is closed (1SWE, chain D), Ser⁴⁵ now forms a hydrogen bond to biotin.

was similar to that of D4, consistent with the least stable subunit dominating the stability of the tetramer.

DISCUSSION

SA's binding to biotin is one of the strongest known non-covalent protein–ligand interactions and yet the mutant protein Tr has 10-fold greater binding stability [24]. The crystal structures of apo- and biotin-bound Tr described in the present paper show the basis for Tr's tenacious ligand binding. The key differences between Tr and SA are the decreased flexibility of L3/4, the increased stability of the Ser⁴⁵ hydrogen bond to biotin and the reduced conformational change upon biotin binding.

The crystal structure of D128A SA and molecular dynamic simulations [15,36,37] led to a model whereby a water molecule initiates biotin dissociation, by competing with the Asp¹²⁸ hydrogen bond to biotin and so promoting co-operative breakage of other biotin hydrogen bonds. However, we note that in wild-type SA the Asp¹²⁸ hydrogen bond to biotin is always in place, but the Ser⁴⁵ hydrogen bond is often broken [28]. Hence we suggest that breakage of the Ser⁴⁵ hydrogen bond to biotin is frequent and is the first event in biotin dissociation. The mutations that led to a higher stability mutant Tr, further support this revised model: in Tr, the Ser⁴⁵ hydrogen bond to biotin was clearly present, whereas the residue that competes with biotin for hydrogen bonding to Ser⁴⁵, Ser⁵², is mutated to a glycine residue. Arg⁵³ does not have direct interactions with residues binding to biotin, so the R53D

mutation is likely to exert its effects by further reducing L3/4 mobility.

L3/4 is always in a 'closed' conformation in biotin-bound structures of SA, acting like a lid over the binding pocket and contributing to the low rate of biotin dissociation. For avidin, L3/4 is ordered when bound to free biotin, but disordered when bound to biotin conjugates: this explains why avidin binds biotin conjugates much more weakly than biotin [38]. A restrained L3/4 may also be important in enabling high-affinity biotin binding by the dimeric rhizavidin [39]. In a series of structures solved in the pH range 2.0–3.1 [40], L3/4 was closed in apo-SA, but in the apo-SA structures solved in the more physiologically relevant pH range 4.5–7.5, L3/4 is disordered, in two (1SWC) or three (1SWA and 1SWB) of the subunits of the tetramer [28]. The stability of L3/4 in the closed conformation in apo-Tr indicates that in biotin–Tr the occasional fluctuations that lead to lid opening, and so facilitate biotin dissociation, are similarly suppressed. Since we found that the hydrogen-bond lengths to biotin and conformation of biotin are equivalent in Tr and SA, this indicates that it is not subtle differences in the ground-state binding conformation that explain the change in stability, but a change in the frequency of alternative protein conformations with weakened binding. The L3/4 conformation in apo-Tr also rationalizes the decreased on-rate of Tr [24], since it will be harder for biotin to enter the binding site with the loop shut. For *Streptomyces avidinii*, one can hypothesize that evolution would favour a SA with a high on-rate as well as a low off-rate, but for laboratory applications in anchoring and bridging [24] the off-rate is more important.

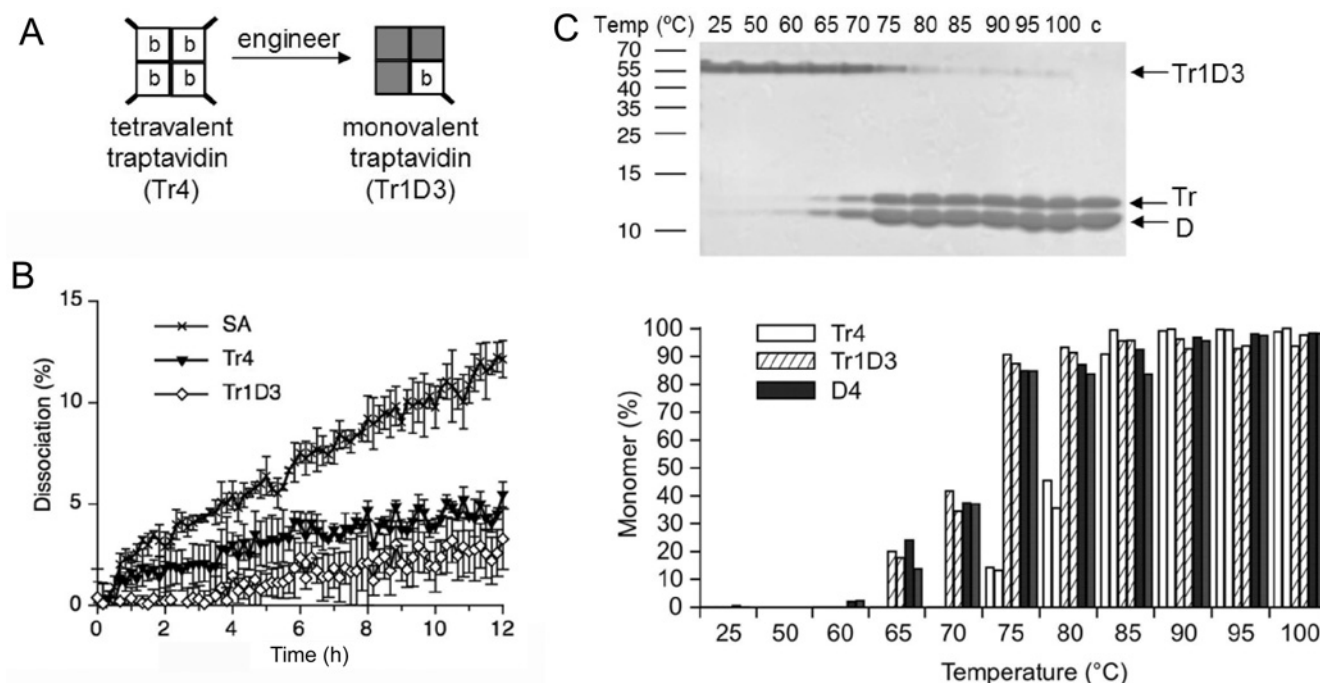


Figure 5 Monovalent Tr off-rate and thermostability

(A) Tr4 (left-hand panel) is tetravalent, with each subunit binding one biotin (b). Monovalent Tr contains one Tr subunit (white square) and three D subunits (grey squares), which cannot bind biotin. Tr but not D subunits have His₆ tags (bold diagonal line). (B) Off-rate of a biotin conjugate from Tr1D3 compared with tetravalent SA and Tr4, in the presence of competing free biotin at 37 °C. Means of triplicate readings are shown $\pm$ S.D. (C) Thermostability of the monovalent traptavidin tetramer, incubated at the indicated temperature for 3 min and analysed by SDS/PAGE and Coomassie Blue staining. The positive control (c) was mixed with SDS before heating at 95 °C. Bands from tetrameric Tr1D3 or monomers (Tr or D) are indicated. The percentage monomer from Tr1D3 from duplicate gels (adjacent striped columns) is plotted against the temperature in the lower panel, with comparison with Tr4 (white columns) or D4 (filled columns).

SA is a highly thermostable protein (72 °C for subunit dissociation without biotin), even though the protein derives from a mesophilic bacterium [1,2,34]. This thermostability relates to reducing the rare subunit-unfolding events that would contribute to ligand dissociation, or, from an energetic perspective, rigid binding sites having a reduced entropic cost for providing an inflexible ligand-binding site [41]. Any fleeting disruption of the interaction between biotin and Trp¹²⁰ of the neighbouring subunit should transiently increase the rate of biotin dissociation 10⁴-fold [42]. Tr is 10 °C more thermostable than SA [24] and one might think that this thermostability would contribute to the increased ligand-binding stability, by further reducing the frequency of conformational fluctuations that disrupt the protein structure or subunit packing. The equivalent off-rate of Tr1D3 and Tr4 demonstrated that the change in Tr's biotin-binding stability is not mediated through altered subunit interactions. The tetravalent nature of Tr4 interferes with many applications: for example in imaging cell-surface proteins, tetravalency can induce receptor cross-linking and activate cell signalling [6,26]. Tr1D3 (and Tr2D2 or Tr3D1, which could be generated, as before [26]) should be valuable in diverse applications, including nanoassembly and the targeting of quantum dots for photostable single-molecule imaging [5,6].

The stability of the interaction between biotin and avidin/SA is paradoxical, given the known binding energy that can be obtained from a given number of hydrogen bonds and from a given surface area, allowing hydrophobic and van der Waals interactions [43,44]. The mutations that generated a higher-stability variant of SA were in the second shell of residues around biotin, pointing to the difficulty of optimizing this kind of high-affinity interaction

computationally. Although we have sufficient knowledge to enable computational and structure-guided design of protein–ligand interactions with micromolar and sometimes nanomolar affinity [44–46], the factors responsible for increasing affinity into the picomolar and femtomolar range are still enigmatic [47,48]. Hydrogen/deuterium-exchange MS, now possible with single-residue resolution, may be able to build upon our crystal structures for further enlightenment on the ‘perfect storm’ of molecular events required for biotin to overcome the large activation barrier to dissociate from Tr [49]. As well as suggesting how to improve further the applications of biotin as an affinity tag, the structures of Tr shed light on the origins of extreme-affinity interactions, with implications for the rational design of drugs with improved efficacy.

ACKNOWLEDGEMENTS

We thank Sonja Baumli for assistance in crystal freezing. Traptavidin is subject to a U.K. priority patent application (0919102.4) filed on 30 October 2009.

FUNDING

Funding was provided by the Wellcome Trust [grant number 085457/Z/08/Z] (A.L.K. and M.H.), the Biotechnology and Biological Sciences Research Council (C.E.C.), Oxford University (E.D.L.) and Worcester College Oxford (M.H.).

REFERENCES

- 1 Sano, T., Vajda, S. and Cantor, C. R. (1998) Genetic engineering of streptavidin, a versatile affinity tag. *J. Chromatogr. B Biomed. Sci. Appl.* **715**, 85–91

- 2 Stayton, P. S., Freitag, S., Klumb, L. A., Chilkoti, A., Chu, V., Penzotti, J. E., To, R., Hyre, D., Le Trong, I., Lybrand, T. P. and Stenkamp, R. E. (1999) Streptavidin-biotin binding energetics. *Biomol. Eng.* **16**, 39–44
- 3 Laitinen, O. H., Hytonen, V. P., Nordlund, H. R. and Kulomaa, M. S. (2006) Genetically engineered avidins and streptavidins. *Cell. Mol. Life Sci.* **63**, 2992–3017
- 4 de Boer, E., Rodriguez, P., Bonte, E., Krijgsvel, J., Katsantoni, E., Heck, A., Grosvel, F. and Strouboulis, J. (2003) Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 7480–7485
- 5 Howarth, M., Takao, K., Hayashi, Y. and Ting, A. Y. (2005) Targeting quantum dots to surface proteins in living cells with biotin ligase. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 7583–7588
- 6 Howarth, M., Liu, W., Puthenveetil, S., Zheng, Y., Marshall, L. F., Schmidt, M. M., Wittrup, K. D., Bawendi, M. and Ting, A. Y. (2008) Monovalent, reduced-size quantum dots for single molecule imaging of receptors in living cells. *Nat. Methods* **5**, 397–399
- 7 Kattah, M. G., Coller, J., Cheung, R. K., Oshidary, N. and Utz, P. J. (2008) HIT: a versatile proteomics platform for multianalyte phenotyping of cytokines, intracellular proteins and surface molecules. *Nat. Med.* **14**, 1284–1289
- 8 Goldenberg, D. M., Sharkey, R. M., Paganelli, G., Barbet, J. and Chatal, J. F. (2006) Antibody pretargeting advances cancer radioimmunodetection and radioimmunotherapy. *J. Clin. Oncol.* **24**, 823–834
- 9 Avrantinis, S. K., Stafford, R. L., Tian, X. and Weiss, G. A. (2002) Dissecting the streptavidin–biotin interaction by phage-displayed shotgun scanning. *ChemBioChem* **3**, 1229–1234
- 10 Aslan, F. M., Yu, Y., Mohr, S. C. and Cantor, C. R. (2005) Engineered single-chain dimeric streptavidins with an unexpected strong preference for biotin-4-fluorescein. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 8507–8512
- 11 Levy, M. and Ellington, A. D. (2008) Directed evolution of streptavidin variants using *in vitro* compartmentalization. *Chem. Biol.* **15**, 979–989
- 12 Green, N. M. (1975) Avidin. *Adv. Protein Chem.* **29**, 85–133
- 13 Wilbur, D. S., Pathare, P. M., Hamlin, D. K., Stayton, P. S., To, R., Klumb, L. A., Buhler, K. R. and Vessella, R. L. (1999) Development of new biotin/streptavidin reagents for pretargeting. *Biomol. Eng.* **16**, 113–118
- 14 Weber, P. C., Pantoliano, M. W., Simons, D. M. and Salemme, F. R. (1994) Structure-based design of synthetic azobenzene ligands for streptavidin. *J. Am. Chem. Soc.* **116**, 2717–2724
- 15 Grubmuller, H., Heymann, B. and Tavan, P. (1996) Ligand binding: molecular mechanics calculation of the streptavidin–biotin rupture force. *Science* **271**, 997–999
- 16 DeChancie, J. and Houk, K. N. (2007) The origins of femtomolar protein–ligand binding: hydrogen-bond cooperativity and desolvation energetics in the biotin–(strept)avidin binding site. *J. Am. Chem. Soc.* **129**, 5419–5429
- 17 Cerutti, D. S., Le, T. I., Stenkamp, R. E. and Lybrand, T. P. (2009) Dynamics of the streptavidin–biotin complex in solution and in its crystal lattice: distinct behavior revealed by molecular simulations. *J. Phys. Chem. B* **113**, 6971–6985
- 18 Miyamoto, S. and Kollman, P. A. (1993) Absolute and relative binding free energy calculations of the interaction of biotin and its analogs with streptavidin using molecular dynamics/free energy perturbation approaches. *Proteins* **16**, 226–245
- 19 Moy, V. T., Florin, E. L. and Gaub, H. E. (1994) Intermolecular forces and energies between ligands and receptors. *Science* **266**, 257–259
- 20 Bruneau, E., Sutter, D., Hume, R. I. and Akaaboune, M. (2005) Identification of nicotinic acetylcholine receptor recycling and its role in maintaining receptor density at the neuromuscular junction *in vivo*. *J. Neurosci.* **25**, 9949–9959
- 21 Dressman, D., Yan, H., Traverso, G., Kinzler, K. W. and Vogelstein, B. (2003) Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 8817–8822
- 22 Swift, J. L., Heuff, R. and Cramb, D. T. (2006) A two-photon excitation fluorescence cross-correlation assay for a model ligand–receptor binding system using quantum dots. *Biophys. J.* **90**, 1396–1410
- 23 Pierres, A., Touchard, D., Benoliel, A. M. and Bongrand, P. (2002) Dissecting streptavidin–biotin interaction with a laminar flow chamber. *Biophys. J.* **82**, 3214–3223
- 23a Howarth, M. and Chivers, C. E. (2009) U. K. Pat. Appl. 0919102.4
- 24 Chivers, C. E., Crozat, E., Chu, C., Moy, V. T., Sherratt, D. J. and Howarth, M. (2010) A streptavidin variant with slower biotin dissociation and increased mechanostability. *Nat. Methods* **7**, 391–393
- 25 Crozat, E., Meglio, A., Allemand, J. F., Chivers, C. E., Howarth, M., Venien-Bryan, C., Grainge, I. and Sherratt, D. J. (2010) Separating speed and ability to displace roadblocks during DNA translocation by FtsK. *EMBO J.* **29**, 1423–1433
- 26 Howarth, M., Chinnappen, D. J., Gerrow, K., Dorrestein, P. C., Grandy, M. R., Kelleher, N. L., El Hussein, A. and Ting, A. Y. (2006) A monovalent streptavidin with a single femtomolar biotin binding site. *Nat. Methods* **3**, 267–273
- 27 Collaborative Computational Project Number 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **50**, 760–763
- 28 Freitag, S., Le Trong, I., Klumb, L., Stayton, P. S. and Stenkamp, R. E. (1997) Structural studies of the streptavidin binding loop. *Protein Sci.* **6**, 1157–1166
- 28b Chen, B., Arendall, III, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 12–21
- 29 Hyre, D. E., Le Trong, I., Merritt, E. A., Eccleston, J. F., Green, N. M., Stenkamp, R. E. and Stayton, P. S. (2006) Cooperative hydrogen bond interactions in the streptavidin–biotin system. *Protein Sci.* **15**, 459–467
- 30 Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Zwart, P. H., Hung, L. W., Read, R. J. and Adams, P. D. (2008) Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **64**, 61–69
- 31 Kada, G., Falk, H. and Gruber, H. J. (1999) Accurate measurement of avidin and streptavidin in crude biofluids with a new, optimized biotin–fluorescein conjugate. *Biochim. Biophys. Acta* **1427**, 33–43
- 32 Jones, M. L. and Kurban, G. P. (1995) Noncooperativity of biotin binding to tetrameric streptavidin. *Biochemistry* **34**, 11750–11756
- 33 Weber, P. C., Ohlendorf, D. H., Wendoloski, J. J. and Salemme, F. R. (1989) Structural origins of high-affinity biotin binding to streptavidin. *Science* **243**, 85–88
- 34 Bayer, E. A., Ehrlich-Rogozinski, S. and Wilchek, M. (1996) Sodium dodecyl sulfate–polyacrylamide gel electrophoretic method for assessing the quaternary state and comparative thermostability of avidin and streptavidin. *Electrophoresis* **17**, 1319–1324
- 35 Li, Q., Gusarov, S., Evoy, S. and Kovalenko, A. (2009) Electronic structure, binding energy, and solvation structure of the streptavidin–biotin supramolecular complex: ONIOM and 3D-RISM study. *J. Phys. Chem. B* **113**, 9958–9967
- 36 Freitag, S., Chu, V., Penzotti, J. E., Klumb, L. A., To, R., Hyre, D., Le Trong, I., Lybrand, T. P., Stenkamp, R. E. and Stayton, P. S. (1999) A structural snapshot of an intermediate on the streptavidin–biotin dissociation pathway. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 8384–8389
- 37 Hyre, D. E., Amon, L. M., Penzotti, J. E., Le Trong, I., Stenkamp, R. E., Lybrand, T. P. and Stayton, P. S. (2002) Early mechanistic events in biotin dissociation from streptavidin. *Nat. Struct. Biol.* **9**, 582–585
- 38 Pazy, Y., Kulik, T., Bayer, E. A., Wilchek, M. and Livnah, O. (2002) Ligand exchange between proteins. Exchange of biotin and biotin derivatives between avidin and streptavidin. *J. Biol. Chem.* **277**, 30892–30900
- 39 Meir, A., Hellepola, S. H., Podoly, E., Nordlund, H. R., Hytonen, V. P., Maatta, J. A., Wilchek, M., Bayer, E. A., Kulomaa, M. S. and Livnah, O. (2009) Crystal structure of rhizavidin: insights into the enigmatic high-affinity interaction of an innate biotin-binding protein dimer. *J. Mol. Biol.* **386**, 379–390
- 40 Katz, B. A. (1997) Binding of biotin to streptavidin stabilizes intersubunit salt bridges between Asp⁶¹ and His⁸⁷ at low pH. *J. Mol. Biol.* **274**, 776–800
- 41 Moghaddam, S., Inoue, Y. and Gilson, M. K. (2009) Host–guest complexes with protein–ligand-like affinities: computational analysis and design. *J. Am. Chem. Soc.* **131**, 4012–4021
- 42 Chilkoti, A., Tan, P. H. and Stayton, P. S. (1995) Site-directed mutagenesis studies of the high-affinity streptavidin–biotin complex: contributions of tryptophan residues 79, 108, and 120. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 1754–1758
- 43 Kuntz, I. D., Chen, K., Sharp, K. A. and Kollman, P. A. (1999) The maximal affinity of ligands. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 9997–10002
- 44 Houk, K. N., Leach, A. G., Kim, S. P. and Zhang, X. (2003) Binding affinities of host–guest, protein–ligand, and protein–transition-state complexes. *Angew. Chem. Int. Ed. Engl.* **42**, 4872–4897
- 45 Das, R. and Baker, D. (2008) Macromolecular modeling with rosetta. *Annu. Rev. Biochem.* **77**, 363–382
- 46 Holm, L., Moody, P. and Howarth, M. (2009) Electrophilic affibodies forming covalent bonds to protein targets. *J. Biol. Chem.* **284**, 32906–32913
- 47 Midelfort, K. S., Hernandez, H. H., Lippow, S. M., Tidor, B., Drennan, C. L. and Wittrup, K. D. (2004) Substantial energetic improvement with minimal structural perturbation in a high affinity mutant antibody. *J. Mol. Biol.* **343**, 685–701
- 48 Foote, J. and Eisen, H. N. (2000) Breaking the affinity ceiling for antibodies and T cell receptors. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 10679–10681
- 49 Rand, K. D., Zehl, M., Jensen, O. N. and Jorgensen, T. J. (2009) Protein hydrogen exchange measured at single-residue resolution by electron transfer dissociation mass spectrometry. *Anal. Chem.* **81**, 5577–5584

Received 6 October 2010/12 January 2011; accepted 18 January 2011

Published as BJ Immediate Publication 18 January 2011, doi:10.1042/BJ20101593

SUPPLEMENTARY ONLINE DATA

How the biotin–streptavidin interaction was made even stronger: investigation via crystallography and a chimaeric tetramer

Claire E. CHIVERS, Apurba L. KONER, Edward D. LOWE and Mark HOWARTH¹

Department of Biochemistry, Oxford University, South Parks Road, Oxford OX1 3QU, U.K.

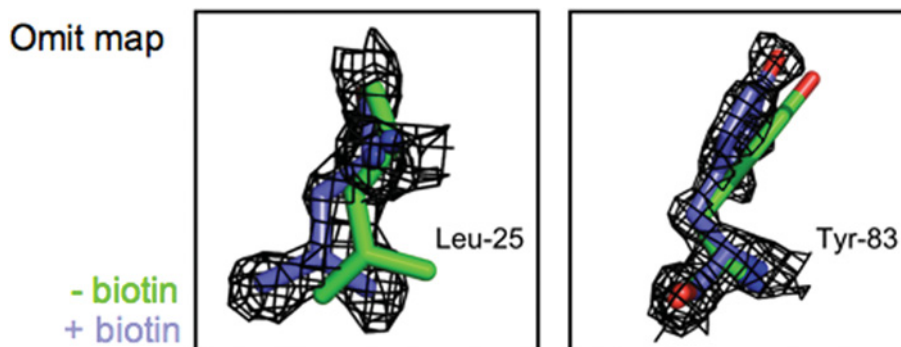


Figure S1 Testing for bias in the biotin–Tr structure caused by the search model for molecular replacement

Apo-Tr was used as the search model for molecular replacement to solve the structure of biotin–Tr. Simulated annealing composite omit maps, with starting phases from the apo-Tr structure, are shown in black mesh for Leu²⁵ (left-hand panel) and Tyr⁸³ (right-hand panel), contoured at 1.3 σ . The apo-Tr backbone structure is overlaid in green and the biotin–Tr structure in blue. The final biotin–Tr structure fitted the omit map much better (correlation coefficient 0.73 for side-chain atoms) than the apo-Tr structure (correlation coefficient 0.47 for side-chain atoms). Leu²⁵ and Tyr⁸³ were chosen because they illustrate that the biotin–Tr structure fits the omit map better than the apo-Tr structure does. Overall, the alignment clearly shows that the apo-Tr model does not fit the composite omit maps, which indicates that the solution of the biotin–Tr structure was not biased by the search model.

¹ To whom correspondence should be addressed (email mark.howarth@bioch.ox.ac.uk).

The structural co-ordinates reported will appear in the PDB under accession codes 2Y3E and 2Y3F.

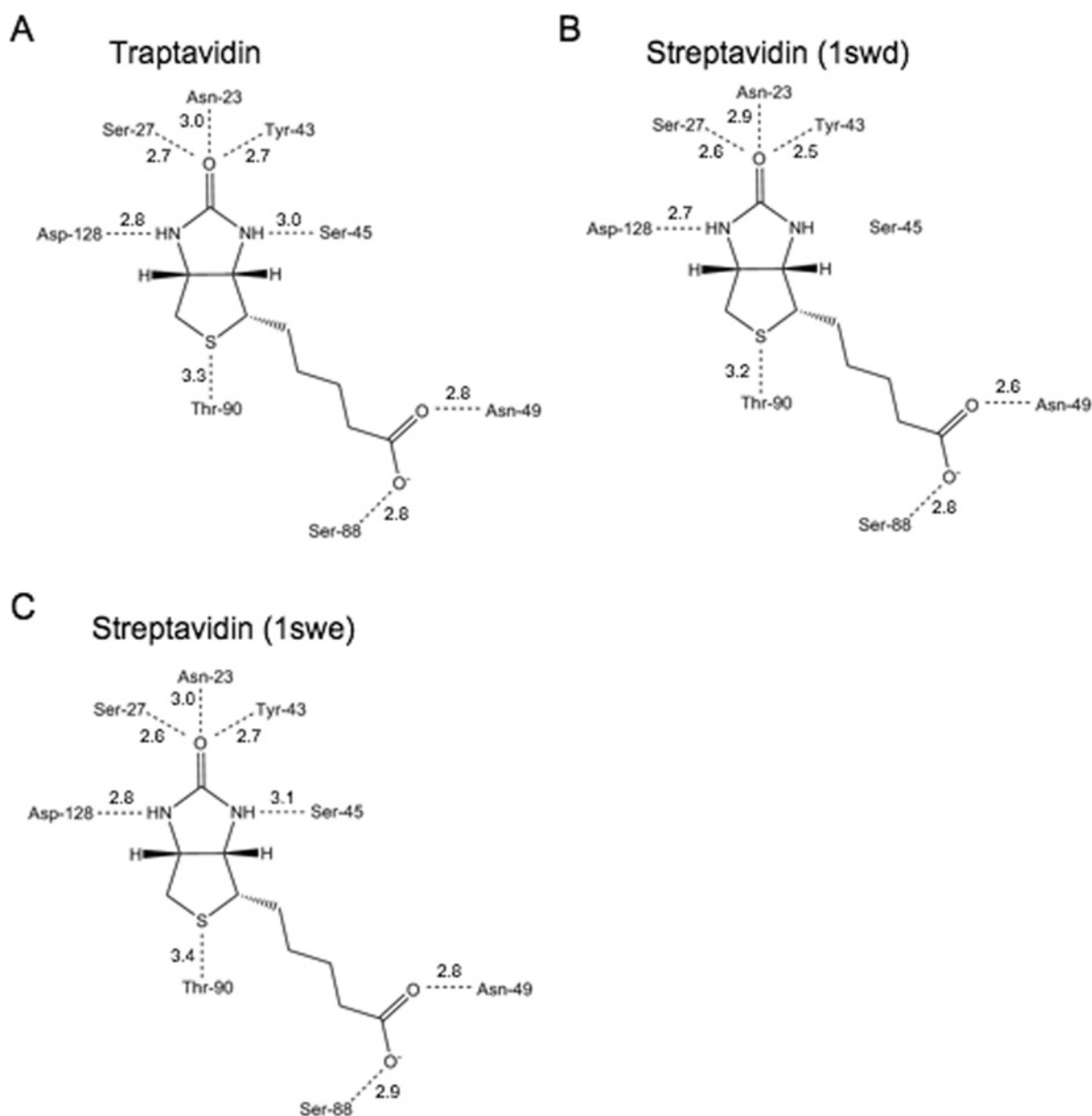


Figure S2 Hydrogen-bond lengths to biotin for Tr and SA

(**A**) Hydrogen bond lengths (in Å) for Tr (only one subunit shown, since all subunits are equivalent). The DPI gave an estimated co-ordinate error for biotin–Tr of 0.051 Å. (**B**) Hydrogen-bond lengths for SA, 1SWD (mean bond lengths for the two biotins bound per tetramer). Ser⁴⁵ here does not form a hydrogen bond to biotin. (**C**) Hydrogen-bond lengths to biotin for SA from an alternative crystal structure, 1SWE (mean bond lengths for the four biotins bound per tetramer).

Received 6 October 2010/12 January 2011; accepted 18 January 2011
Published as BJ Immediate Publication 18 January 2011, doi:10.1042/BJ20101593