

Señora

Edna Marcela Ramírez Orozco

Directora de Nuevas Creaciones

Superintendencia de Industria y Comercio

E. S. D.

Referencia: Solicitud de PCT Patente de Invención denominada ***Proteína Recombinante que se une a HER2 y Seroalbumina.***

Asunto: Respuesta al Requerimiento No. 15205 del 1 de diciembre de 2020.

Expediente No. NC2019/0002609

Juan Carlos Cuesta Quintero, mayor de edad, identificado como aparece al pie de mi firma, obrando en mi calidad de apoderado de la sociedad **Molecular Partners AG**, doy respuesta al Oficio No. 15205 del 1 de diciembre de 2020, en los siguientes términos:

1. Título:

De acuerdo con las objeciones propuestas por parte del Examinador en cuanto al título de la invención, la Solicitante acepta el título propuesto por parte de este Despacho.

En ese sentido, solicitamos respetuosamente se tenga como título de la invención el siguiente:

Proteína Recombinante que se une a HER2 y Seroalbumina

2. Reivindicaciones:

- La nueva reivindicación 1 ha sido modificada para aumentar la identidad de secuencia requerida con la SEQ ID NO: 21 a al menos un 98%. El soporte para esta enmienda se puede encontrar en la especificación, p.ej. en la página 24, líneas 13 a 25 y página 25 líneas 1 a 6, del texto de la solicitud originalmente presentado.
- La nueva reivindicación 2 se ha modificado para aumentar la identidad de secuencia requerida con la SEQ ID NO: 21 a al menos un 99%. El soporte para esta enmienda se puede encontrar en la especificación, p.ej. en la página 24, líneas 13 a 25, del texto de la solicitud originalmente presentado.
- Las reivindicaciones 9 a 13 se han cancelado. El resto de reivindicaciones se han reenumerado y su dependencia se ha ajustado en consonancia.

- Las reivindicaciones 1, 2, 11, 13 no son claras ni tienen sustento porque la expresión “al menos” para referirse a porcentajes de identidad del 92%, 95, 90% o 98%, no permiten definir el alcance de la invención. Toda vez que, la invención incluiría secuencias que no presentan las propiedades mejoradas que demostró la SEQ ID NO: 21 (Ej. SEQ ID NO: 22 y 23). Se sugiere definir las secuencias diseñadas y probadas en el capítulo reivindicatorio por su SEQ ID NO.

En primer lugar, el Solicitante desea mencionar que las nuevas reivindicaciones modificadas adjuntas se refieren ahora a proteínas de unión recombinante que comprenden una secuencia de aminoácidos que tiene al menos un 98 % o 99% (reivindicaciones 1 y 2 respectivamente) de identidad de secuencia de aminoácidos con la SEQ ID NO: 21.

En segundo lugar, el Solicitante discrepa respetuosamente con la objeción por falta de claridad de la expresión “al menos” para referirse a porcentajes de identidad del 98% o 99% en relación a que no permiten definir el alcance de la invención. El término “al menos” en ningún caso impreciso. En contra de la opinión de la Examinadora, este término, no es impreciso, porque permite una comparación con el estado de la técnica de forma que por ejemplo un documento que divulgase la característica de “97% de identidad con la SEQ ID NO: 21”, se podría comparar con la materia-objeto de las reivindicaciones y claramente no caería dentro de la característica “al menos un 98% de identidad con la SEQ ID NO: 21”. Cualquier otro porcentaje de identidad con la SEQ ID NO: 21, por ejemplo, 98%, 99% o incluso 100%, sí caerían dentro de dicha característica y la anticiparían. Por ello este término es claro, ya que sí permite la comparación con el estado de la técnica y delimitar lo que se protege frente al estado de la técnica de forma clara y precisa.

De hecho, el término “al menos” no es un término como sería “aproximadamente” que en efecto introduce una cierta imprecisión de los límites del valor correspondiente (ver la guía de Examen, particularmente el punto 2.6.5.3). Es más, dicho término “al menos” aparece en multitud de patentes concedidas en numerosos países, incluidos Colombia. Así, por ejemplo, las patentes concedidas No. 14-065.568; No. 13-126.418, No. 14-132613 y No.14-132629, entre otras, comprenden reivindicaciones en las que el término “al menos” aparece, incluso varias veces.

Así, teniendo en cuenta lo anteriormente mencionado, el Solicitante, respetuosamente considera que el término “al menos” presenta claridad y, por lo tanto, la presente objeción ha quedado superada.

En cuanto a las propiedades mejoradas de la proteína de unión reivindicada que presenta al menos un 98% de identidad con la SEQ ID NO: 21, el Solicitante respetuosamente dirige a la Examinadora a la discusión del soporte en la

descripción y actividad inventiva que aparecen a continuación en la presente respuesta.

Así, teniendo en cuenta lo anteriormente mencionado, junto con la información comprendida en el apartado de actividad inventiva y soporte en la descripción que aparecen más adelante, el Solicitante respetuosamente considera que las nuevas reivindicaciones 1 y 2 presentan claridad y están suficientemente soportadas por la información y ejemplos incluidos en la presente solicitud de patente. Por lo tanto, la presente objeción no aplica.

- Las reivindicaciones 1, 9 y 12 no son concisas porque reclaman una proteína de unión recombinante en más de una reivindicación independiente.

Para solventar la presente objeción el Solicitante ha eliminado las reivindicaciones 9 y 12, por lo que ahora, el nuevo set de reivindicaciones modificado que acompaña al presente escrito presenta una única reivindicación independiente, la reivindicación 1, que reivindica la proteína de unión recombinante de la invención.

Así, teniendo en cuenta lo anteriormente mencionado, el Solicitante, respetuosamente considera que la presente objeción ha quedado superada.

- La reivindicación 13 no es clara porque se refiere a cualquiera de las reivindicaciones 1 a 12 y menciona que la proteína tiene al menos 98% de identidad con la SEQ ID NO: 21, pero la reivindicación 12 reclama una proteína que comprende la SEQ ID NO: 21, por lo que el alcance de la reivindicación 13 es más amplio que el de la reivindicación de la cual depende. Se sugiere realizar las modificaciones pertinentes.

La reivindicación 13 ha sido eliminado del nuevo set de reivindicaciones modificadas que acompaña al presente escrito. Por lo tanto, el Solicitante respetuosamente considera que la presente objeción ha quedado superada.

Las modificaciones descritas anteriormente no se extienden más allá del contenido de la solicitud tal como fue presentada; tampoco resultan en una ampliación del alcance de la materia originalmente reivindicada.

3.Descripción:

La Examinadora objeta que la descripción no divulga la invención de manera suficientemente completa para que la persona normalmente versada en la materia técnica correspondiente pueda ejecutarla (o reproducirla). Toda vez que, no se evidencia el

desarrollo ni pruebas técnicas en el capítulo descriptivo de variantes de la SEQ ID NO: 21 que correspondan con 90%, 92%, 95% o 98% de identidad.

El Solicitante discrepa respetuosamente con esta objeción.

Primero, el Solicitante señala que, sin aceptar la objeción y en un esfuerzo por acelerar la concesión de esta solicitud, se han modificado las reivindicaciones 1 y 2 para aumentar el requisito de identidad de secuencia de las secuencias de aminoácidos hasta al menos un 98% o 99%, respectivamente, de identidad de secuencia con la SEQ ID NO: 21.

Con respecto a la falta de evidencia del desarrollo ni pruebas técnicas en el capítulo descriptivo de variantes de la SEQ ID NO: 21 que correspondan con el porcentaje de identidad reivindicados, el Solicitante afirma respetuosamente que la especificación, combinada con el conocimiento en la técnica en el momento de la presentación de la presente solicitud de patente, brinda suficiente respaldo para las reivindicaciones enmendadas. La relación estructura-función general de los dominios diseñados con repeticiones de anquirina era bien conocida en la técnica en el momento de su presentación. Como se describe en la especificación originalmente presentada (ver, por ejemplo, página 53, líneas 18-25) y como se conoce en la técnica previa (ver, por ejemplo, Binz et al., 2004 (ver abajo y adjunto)), un dominio diseñado con repeticiones de anquirina típicamente comprende un módulo N-terminal, dos o tres módulos internos de repeticiones de anquirina y un módulo C-terminal (es decir, una estructura "N2C" o "N3C").

Dentro de esta estructura N2C o N3C, los posibles sitios de interacción con el objetivo (por ejemplo, la interacción mediadora con HER2 o seroalbúmina) se encuentran en los módulos internos con repeticiones de anquirina. Los módulos de los extremos N-terminal y C-terminal son repeticiones terminales especiales diseñados para proteger los módulos internos de repetición de anquirina que contienen residuos de potencial interacción con el objetivo.

La ubicación de los posibles sitios de interacciones con el objetivo dentro de la estructura de los dominios con repeticiones de anquirina también se conocían. Por ejemplo, Binz et al., 2004 (adjunto), proporcionaron una explicación detallada de la estructura general de un dominio diseñado con repeticiones de anquirina y la ubicación de los potenciales sitios de interacción con el objetivo dentro de la secuencia del módulo con repeticiones de anquirina:

*“Diseñamos un módulo AR de consenso que consta de **seis potenciales residuos de interacción diversificados** (que pueden ser cualquier aminoácido excepto cisteína, glicina y prolina) y 27 residuos marco (26 son fijos y uno puede ser asparagina, histidina o tirosina). Este módulo fue diseñado a partir de alineamientos de secuencia y análisis estructurales (Figura 1). **Los posibles residuos de interacción aleatorios se***

encuentran en el giro β y la primera hélice α del módulo AR. Clonamos cantidades variables de estos módulos de repeticiones entre las repeticiones terminales, que son repeticiones terminales especiales de dominios AR que protegen el núcleo hidrofóbico (Fig. 1)."

(Fragmento traducido de Binz et al., Nature Biotech. 22(5), p.575-582 (2004), en la p.575, 2ª col., y p.576, 1ª col. (énfasis añadido, copia adjunta)).

Por lo tanto, los módulos internos con repeticiones de anquirina de un dominio con repeticiones de anquirina proporcionan la función de unión específica con el objetivo, mientras que la función principal de los módulos terminales de protección es estabilizar el dominio de repetición de anquirina y proteger su núcleo hidrofóbico proporcionado por los módulos internos de repetición de anquirina. Estas funciones separadas de los módulos terminales de protección y los módulos internos con repeticiones de anquirina de los dominios diseñados con repeticiones de anquirina se ilustran muy bien, por ejemplo, en las Figuras 1c y 4a de Binz et al., 2004, que muestran la estructura 3D de un dominio de repetición de anquirina, ya sea solo o en un complejo con su proteína objetivo. Los posibles residuos de interacción objetivo se muestran en rojo y están ubicados en los módulos internos con repeticiones de anquirina.

Otros documentos del estado de la técnica anterior proporcionaron información detallada adicional sobre la estructura 3D de los dominios con repeticiones de anquirina, su interacción molecular con proteínas objetivo y la ubicación de potenciales residuos de interacción y residuos de esqueleto dentro de la estructura 3D (ver, por ejemplo, Sennhauser y Grütter, "Chaperone-Assisted Crystallography with DARPins", Structure, vol. 16, p.1443-53 (2008); copia adjunta).

Con respecto a la variabilidad de secuencia limitada permitida en las reivindicaciones modificadas, el Solicitante sostiene que un experto en la materia sabía de las revelaciones en la descripción de la invención y en la técnica previa, p.ej. Binz et al., 2004, y Sennhauser y Grütter, 2008, donde en las repeticiones de anquirina recitadas se pueden introducir cambios de aminoácidos en los dominios sin afectar la estructura general de los dominios con repeticiones de anquirina y sin cambiar la especificidad de unión para HER2 o seroalbúmina.

Además, la especificación y la técnica anterior describen métodos de rutina para probar si una proteína con repeticiones de anquirina se une a un objetivo específico, como HER2 o seroalbúmina, incluyendo ensayos ELISA y análisis de resonancia del plasmón de superficie (véase la especificación, por ejemplo, el Ejemplo 5, páginas 70-71 de la solicitud originalmente presentada). Por lo tanto, un experto en la materia podría determinar sin indebida experimentación si un dominio con repeticiones de anquirina se une

específicamente a HER2 o seroalbúmina, y cae dentro del ámbito de protección de las nuevas reivindicaciones enmendadas.

En vista de todo lo anterior, el Solicitante afirma respetuosamente que la descripción de la invención, en combinación con el conocimiento en el arte previo en el momento de su presentación, proporciona suficiencia técnica y claridad a las nuevas reivindicaciones enmendadas.

4. Examen de Patentabilidad:

4.1. Nivel Inventivo:

La Examinadora objeta las reivindicaciones originales 1-16 por falta de nivel inventivo a la luz de la información descrita en **D1** (WO2014/083208).

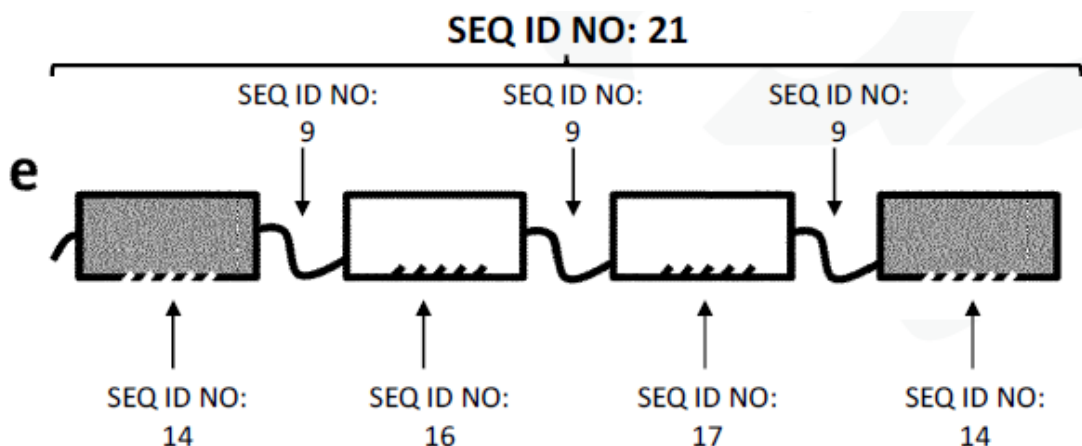
Brevemente, la Examinadora indica que no se evidencia un efecto técnico asociado a la diferencia técnica entre la presente invención y **D1**, es decir frente una secuencia de aminoácidos que tiene al menos 92% de identidad de secuencia de aminoácidos con la SEQ ID NO: 21 y la proteína de unión recombinante de **D1**. Toda vez que, no se evidencia el desarrollo ni pruebas técnicas asociadas a variantes con al menos 92% de identidad de secuencia con la SEQ ID NO: 21 de la invención. Teniendo en cuenta que **D1** ya enseña el diseño de proteínas recombinantes que comprenden dominios de anquirina que se unen a HER2, a modo de ejemplo, las SEQ ID NO: 62-68, 72 (Pág. 58, línea [2], Reiv. 12, listado de secuencias), y además, dichas proteínas recombinantes comprenden dominios de anquirina con especificidad por seroalbúmina (Pág. 45, línea [13]), un experto en la materia se vería motivado a incluir una variante de secuencia con dominios de anquirina que se unen a HER2 y con dominios de anquirina con especificidad para seroalbúmina del documento **D1** en la proteína recombinante, para así llegar al objeto de la reivindicación 1. Por dichos motivos, la Examinadora considera que la presente invención carece de actividad inventiva.

Las modificaciones llevadas a cabo en las nuevas reivindicaciones 1 y 2 han limitado en mayor medida el objeto de la presente invención y ahora hacen referencia a una proteína de unión recombinante que comprende una secuencia de aminoácidos que tiene al menos un 98 % o 99% de identidad de secuencia de aminoácidos con la SEQ ID NO: 21, en donde dicha proteína de unión recombinante comprende dos dominios diseñados con repeticiones de anquirina con especificidad de unión para HER2 y dos dominios diseñados con repeticiones de anquirina con especificidad de unión por la seroalbúmina, y donde dicha proteína de unión recombinante presenta unas ventajas y efectos técnicos sorprendentes a la luz del estado de la técnica citado, **D1**.

A continuación se exponen los argumentos a favor de la actividad inventiva del nuevo conjunto de reivindicaciones adjunto a la luz del estado de **D1**.

La SEQ ID NO: 21 codifica para una proteína de unión recombinante que comprende dos dominios diseñados con repeticiones de anquirina con especificidad de unión para HER2 (SEQ ID NO: 16 y 17), que están flanqueados por dos dominios diseñados con repeticiones de anquirina con especificidad de unión por la seroalbúmina (SEQ ID NO: 14). Dichos cuatro dominios diseñados con repeticiones de anquirina están unidos a través de enlaces peptídicos ricos en Pro-Thr (SEQ ID NO: 9).

La estructura de la SEQ ID NO: 21 se muestra en la **Figura 1 (e)** (véase también la descripción en la página 5, líneas 20-25 y página 6, líneas 5-11) de la presente solicitud, que se muestra a continuación para comodidad de la Examinadora.



La Examinadora alega que no se ha demostrado ningún efecto técnico inventivo asociado a una proteína de unión recombinante que tiene al menos 92% de identidad de secuencia de aminoácidos con la SEQ ID NO: 21.

El Solicitante está respetuosamente en desacuerdo, porque, como se discute en la descripción de la presente invención, los inventores han encontrado efectos inesperados y sorprendentes para dicha proteína reivindicada.

Los inventores han encontrado sorprendentemente que al tener dos dominios diseñados con repeticiones de anquirina con especificidad de unión para la albúmina sérica en lugar de uno en la proteína de unión recombinante, las propiedades farmacocinéticas de la proteína de unión recombinante pueden mejorarse (ver página 2, líneas 19-23 de la solicitud originalmente presentada).

El solicitante remite respetuosamente a la Examinadora a la página 26, líneas 16-25 y siguientes, de la solicitud originalmente presentada, donde se indica lo siguiente:

“El dominio diseñado con repeticiones de anquirina de la SEQ ID NO: 14 presenta propiedades de estabilidad en almacenamiento mejorada

(véase el Ejemplo 2 y 3; Figura 2) en comparación con los dominios diseñados con repeticiones de anquirina conocidos con especificidad de unión para la seroalbúmina. Dichos dos dominios diseñados con repeticiones de anquirina con especificidad de unión para la seroalbúmina dan lugar sorprendentemente a propiedades farmacocinéticas mejoradas de la proteína de unión recombinante en comparación con una proteína de unión recombinante correspondiente que tiene solo un dominio diseñado con repeticiones de anquirina con especificidad de unión para la seroalbúmina (Ejemplo 11, Figura 5)."

Específicamente, en la solicitud originalmente presentada se menciona y demuestra que la proteína #21 (que es una proteína de unión recombinante que consiste en la SEQ ID NO: 21) exhibe propiedades farmacocinéticas notablemente mejoradas, en particular una vida media terminal más larga (**Ejemplo 11; Figura 5**).

Además, los inventores han encontrado sorprendentemente que los enlazadores ricos en Pro-Thr tienen un impacto positivo en las propiedades farmacocinéticas, concretamente una vida media más larga, para una proteína de unión recombinante que comprende dos dominios diseñados con repeticiones de anquirina con especificidad de unión para HER2 y dos dominios diseñados con repeticiones de anquirina con especificidad de unión para la albúmina sérica, en comparación con un enlazador rico en Gly-Ser (**Ejemplos 10 y 11**).

Sorprendentemente, los inventores también descubrieron que la disposición de los dominios diseñados con repeticiones de anquirina utilizados en la Proteína #21 conduce a una mayor expresión recombinante (**véase el Ejemplo 13**).

El Solicitante, respetuosamente afirma que los inventores en la presente solicitud de patente han demostrado, como se ha expuesto anteriormente, efectos inesperados y sorprendentes para las proteínas de unión recombinantes que comprenden la SEQ ID NO: 21.

Así, además de todas las ventajas y efectos técnicos sorprendentes mencionados anteriormente, la presente solicitud de patente también demuestra claramente que las proteínas de unión reivindicadas se unen a HER2 con alta afinidad (**Ejemplo 5**), inhiben la fosforilación de HER2 en células tumorales que expresan HER2 (**Ejemplo 16**), inhiben la proliferación de células tumorales que expresan HER2 (**Ejemplos 6 y 17**), inducen la apoptosis en células tumorales que expresan HER2 (**Ejemplo 6**), e inhiben el crecimiento tumoral en modelos de xenoinjerto de ratón que expresan HER2 (**Ejemplo 7**).

Al menos por todas las razones expuestas anteriormente, el Solicitante respetuosamente considera que el conjunto de reivindicaciones modificado es inventivo a la luz del estado de la técnica citado.



Finalmente, la presente solicitud de patente sigue el procedimiento PPH basado en las reivindicaciones concedidas para la solicitud hermana en Estados Unidos, **US 10,717,772 B2**, la cual ha sido concedida con el mismo alcance de protección que el que presentan las reivindicaciones actualmente en el expediente. Para acelerar la tramitación y posterior concesión de la presente solicitud de patente, el Solicitante procede a llevar a cabo una serie de modificaciones en las reivindicaciones para cumplir con los requisitos exigidos por la presente Oficina.

En vista de todo lo anterior, ruego a usted estudiar nuevamente la solicitud con base en las modificaciones realizadas y proceder de conformidad.

Cordialmente,

Juan Carlos Cuesta Quintero.

C.C. 79.339.809 de Bogotá

T.P. 43.273 del C.S. de la J.

El futuro
es de todosGobierno
de Colombia

BANDEJA DE ENTRADA >

NC2021/0004647 - NC Modificación a las reivindicaciones - Solicitado en Patente NC2019/0002609

Salir

Datos de la solicitud

Número de solicitud	NC2021/0004647
Referencia	P190915CO / FN00620815
Fecha de radicación	14 abr. 2021
Estado	Presentada

Contacto

Apoderado	Número de identificación	Nombre(s)	Apellido(s)	Dirección (es)			
	79339809	JUAN CARLOS	CUESTA QUINTERO	Dirección Física : CLL 72 # 10 -07 OFC 603 BOGOTÁ D.C. (CO)			
Solicitante	Número de identificación	Nombre(s)	Apellido(s)	Dirección (es)			
	1113487		MOLECULAR PARTNERS AG	Dirección Física : Wagistrasse 14 8952 Zürich-Schlieren 8952 (CH)			
Contacto de la solicitud	Número de identificación	Nombre	Dirección	Ciudad	Código postal	País	Tipo de dirección
	79339809	JUAN CARLOS CUESTA QUINTERO	CLL 72 # 10 -07 OFC 603	BOGOTÁ		CO	Dirección Física

1 Solicitud (es)

Número de solicitud	Tipo de solicitud	Tipo de enlace	Título/Denominación	Fecha de radicación	Estado(s)	Titular
---------------------	-------------------	----------------	---------------------	---------------------	-----------	---------

Cerrar

14/4/2021

NC2021/0004647 - NC Modificación a las reivindicaciones - Solicitado en Patente NC2019/0002609

NC2019/0002609	Patente PCT	Vinculado	PROTEÍNAS DE UNIÓN RECOMBINANTES Y SUS USOS	21 mar. 2019	Requerimiento de Examen de Fondo	MOLECULAR PARTNERS AG
--------------------------------	-------------	-----------	---	--------------	--	-----------------------------

Cra 13 No. 27-00 pisos 1, 3, 4, 5, 6, 7 y 10 PBX: (571)5870000 . Call center: (571)5920400 Línea gratuita nacional 018000-910165
www.sic.gov.co . e-mail: contactenos@sic.gov.co . Bogotá D.C. - Colombia.

[Política de privacidad](#) | Política editorial | Créditos | Webmaster: contactenos@sic.gov.co ::: Todos los derechos reservados 2008 - 2021



SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO

NIT: 800.176.089-2

COMPROBANTE DE RECAUDO SISTEMA SIPI

Radicado: NC2021/0004647
Recibo: 21-39527
Fecha: 14/04/2021

JUAN CARLOS CUESTA QUINTERO
CC: 79.339.809
Pago PSE No: 523942
CUS: 956018813

NC Modificación a las reivindicaciones
Request on Patente PCT NC2019/0002609

Detalle del pago

Id.	Rentístico	Concepto y Servicio	Cant.	Val. Unit.	Subtotal
1	50005-01-03	Modificaciones / Correcciones al registro Patente PCT (DESCUENTO INTERNET)	1	\$ 171,500	\$ 171,500
TOTAL					\$ 171,500

Chaperone-Assisted Crystallography with DARPins

Gaby Sennhauser¹ and Markus G. Grütter^{1,*}

¹Department of Biochemistry, University of Zürich, CH-8057 Zürich, Switzerland

*Correspondence: gruetter@bioc.uzh.ch

DOI 10.1016/j.str.2008.08.010

The structure of proteins that are difficult to crystallize can often be solved by forming a noncovalent complex with a helper protein—a crystallization “chaperone.” Although several such applications have been described to date, their handling usually is still very laborious. A valuable addition to the present repertoire of binding proteins is the recently developed designed ankyrin repeat protein (DARPin) technology. DARPins are built based on the natural ankyrin repeat protein fold with randomized surface residue positions allowing specific binding to virtually any target protein. The broad potential of these binding proteins for X-ray crystallography is illustrated by five cocrystal structures that have been determined recently comprising target proteins from distinct families, namely a sugar binding protein, two kinases, a caspase, and a membrane protein. This article reviews the opportunities of this technology for structural biology and the structural aspects of the DARPin-protein complexes.

Despite the continuous technical advances in protein crystallography due to the worldwide effort in structural genomics programs that promote the application of automated procedures in cloning, expression, purification, crystallization, data collection using synchrotron radiation, and computational crystallography, the determination of some protein structures remains a difficult task. Today, the main limitation is the growth of well-diffracting crystals—a pure trial-and-error process—because the formation of protein crystals depends on several unpredictable variables. In particular, advances in microscaling methodology involving pipetting robots and automated imaging of individual experiments have allowed high throughput approaches boosting the field. However, even with these advances, many proteins resist forming suitable crystals on their own because of inherent structural flexibility and instability. Consequently, conformation stabilizing methods that help to crystallize proteins seem attractive.

Approaches in this direction are removal or mutation of surface residues known to be flexible (Lawson et al., 1991), the trimming of the protein's often flexible N- and C-terminal ends, or the analysis of stable domains of larger multidomain proteins (Dere-wenda, 2004). Another successful approach to promote the crystallization of a particular protein has been the addition of a conformation-stabilizing ligand. Current crystallization tools include the building of complexes with diverse known natural or synthetic ligands, such as substrates, inhibitors, nucleic acid, cofactors, or protein ligands. The benefits of such noncovalent complexes can be manifold. A specific conformation of the target protein may be stabilized in such a complex. This increases the chances to obtain crystals of proteins that otherwise adopt several conformations and therefore resist forming crystals. In addition, novel surfaces can potentially provide better crystal contacts. Especially for membrane proteins, protein ligands also increase the hydrophilic surface area of an otherwise small non-transmembrane region. This strategy for the crystallization of membrane proteins has repeatedly led to successful structure determinations, as seen for many antibody fragment-membrane protein complexes (Dutzler et al., 2003; Hunte et al., 2000; Iwata et al., 1995; Zhou et al., 2001). Cocrystallization of a target pro-

tein in complex with a proteinaceous ligand of known structure potentially provides another advantage, namely the structure solution by molecular replacement allowing to avoid the often time-consuming determination of phases via heavy atoms as in single/multiple anomalous dispersion or multiple isomorphous replacement techniques.

DNA and RNA binding proteins are often cocrystallized with their natural binding partner, fragments of DNA or RNA stabilizing the conformation of the protein (Anderson et al., 1984). In many cases, however, the protein of interest lacks a natural binding partner, the natural binding partner might not be known, or it might not be available for other reasons. Therefore, proteinaceous ligands, either natural or synthetic, have received a great deal of attention in the field of structural biology. Antibody and fragments thereof are the most successful molecules to be used as high-affinity and specific binding proteins in biomedical research and have not only been used as a binding protein but also assist in crystallization (Amit et al., 1986). In 1995, Ostermeier et al. selected monoclonal antibodies for the first time against a membrane protein, namely cytochrome c oxidase of *Paracoccus denitrificans*, finally allowing the successful structure determination of the oxidase in complex with a FV fragment (Iwata et al., 1995; Ostermeier et al., 1995). In addition to the well-known Fab and Fv fragments, single-chain antibody domains such as camelid VHH fragments or scFvs led to successful structure determinations (Desmyter et al., 1996; Kortt et al., 1994). Recently, several studies using nonantibody scaffolds such as affibodies (Hogbom et al., 2003; Wahlberg et al., 2003), monobodies (Koide et al., 2007), and repeat proteins (Binz et al., 2004) have shown that these molecules are potential alternatives to antibody fragments in structural biology. These binding molecules were developed as tools for basic research, but their potential use in commercial applications (as specific reagents in biomedical research and as potential therapeutic molecules) was immediately apparent and quickly recognized (Gill and Damle, 2006). Many novel binding proteins have improved on the limitations of antibodies, in particular their rather low production yield and intrinsic stability. Simultaneously, they retain the beneficial properties of antibodies, namely specific and tight

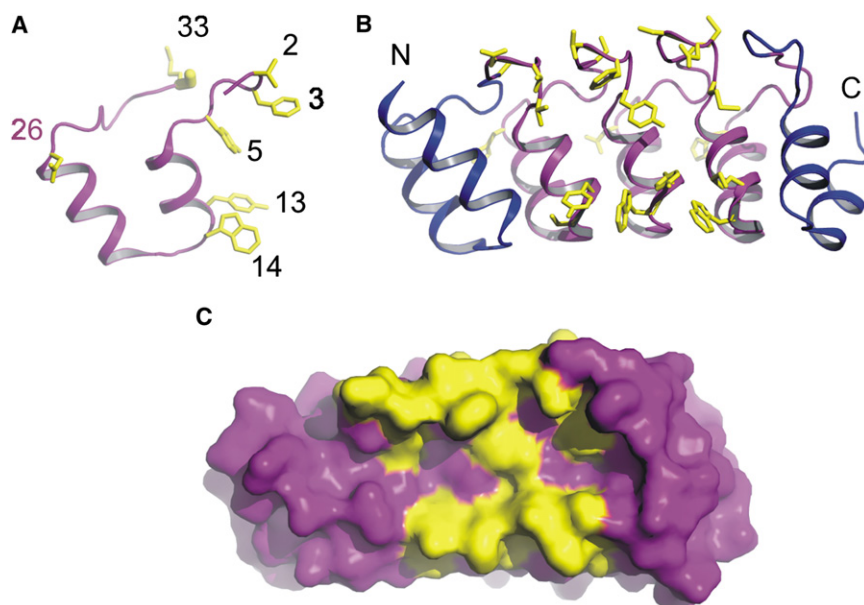


Figure 1. Structure and Potential Interaction Surface of a DARPin

(A) Ribbon representation of a designed ankyrin repeat module. The side chains of the six randomized positions and the randomized framework residue (position 26) are shown as yellow sticks, and their positions within the 33-amino acid conserved framework are labeled. Side chains of the conserved residues are not shown for clarity.

(B) An N3C DARPin library member. The three repeat modules are colored purple; the capping repeats are colored blue. The randomized positions, located mainly in the β turns and the first α helices of the individual repeat modules, are shown as yellow sticks.

(C) Randomized potential interaction surface of an N3C DARPin that normally provides the interaction site (in yellow). The orientation is the same as in (B). All figures were prepared with PyMol (DeLano, 2002).

binding. Structural biologists likewise have a great interest in using these molecules as chaperones in crystallization.

In this review, we focus on one scaffold, the DARPins. The potential of DARPins as promising compounds in drug discovery and drug development has been reviewed recently (Stumpp and Amstutz, 2007). We examine the five recently determined X-ray structures of DARPin complexes and address several questions concerning DARPin-target interactions. The diversity of the successfully chosen antigens is paralleled by the diversity of the interaction partners of natural ankyrin repeat proteins, which indicates that the technology can be generally applied to a very broad range of target molecules.

Drawbacks of Antibodies

The main advantage of antibodies is their ability to recognize a wide range of target molecules with high affinity and specificity. Besides proteins, antibodies can bind compounds such as peptides, sugars and small molecules. However, natural antibodies are complex multidomain molecules not well suited for cocrystallization due to their flexible linker regions between domains and their bivalent character. Therefore, only antibody fragments have been applied for cocrystallization purposes. Fab or Fv fragments contain the complete antigen recognition sites and are therefore sufficient to retain the specific antibody-antigen interaction. Nevertheless, these fragments often derive from monoclonal antibodies, and their classical production by hybridoma technology is time-consuming, expensive, and based on the use of living organisms. The selection using recombinant and synthetic libraries have facilitated their generation (Rothe et al., 2008) and led to successful structure determinations (Fellouse et al., 2007; Fellouse et al., 2005; Fellouse et al., 2004; Ye et al., 2008) but the complicated multidomain molecular architecture with disulfide bonds still complicates their handling, limits their use in reducing environments and requires special production precautions in bacterial expression systems such as expression in the periplasm to guarantee the correct disulfide

bond formation. This often affects the production yield since the volume of the periplasm is limited. Despite these difficulties there are numerous structures in complex with soluble target proteins (Bentley, 1996) as well as more demanding membrane proteins (Hunte and Michel, 2002) but the problems associated with all antibody-derived crystallization chaperones inspire continuously to search for alternative binding proteins with similar binding capabilities but superior properties regarding structure, stability and throughput.

What Are DARPins?

DARPins are derived from the ankyrin repeat motif present in numerous naturally occurring proteins. In nature, ankyrin repeat containing proteins are involved in a wide variety of biological activities and are present in all three superkingdoms (Bork, 1993). Best documented is their involvement in specific protein-protein interactions. The diversity of their roles in a cellular context is further reflected in their localization, which can be the nucleus, cytoplasm, and the extracellular space, where these proteins interact with a large diversity of partners. The number of repeats presented in a single ankyrin repeat protein and thus involved in binding is highly variable so that ankyrin domains can bind to host target molecules that vary considerably in size and shape.

The 33-residue sequence motif of an ankyrin repeat forms a well-defined architecture, consisting of a β turn, followed by a pair of antiparallel α helices and a loop that builds the connection to the next repeat (Figure 1A). The characteristic secondary structure components fold into an L-shaped conformation where the helices form the stem and the loop projects outward at an angle of about 90° . Not uncommon are insertions between or within repeats, mainly in the β turn/loop region (Sedgwick and Smerdon, 1999). These insertions can be either a short helical segment or more complex motifs. The fully assembled ankyrin domain is elongated and slightly curved, manifested particularly in high repeat number ankyrin repeat proteins (Figure 1B). In

naturally occurring complexes involving ankyrin repeat proteins, the concave surface, formed by the β turn and the first α helix, is commonly involved in binding the target molecules as shown by the X-ray structures determined to date: p16^{INK4a}-CDK6 (Russo et al., 1998), p19^{INK4d}-CDK6 (Brotherton et al., 1998; Russo et al., 1998), GABP α -GABP β -DNA (Batchelor et al., 1998), I κ B α -NF κ B (Huxford et al., 1998; Jacobs and Harrison, 1998; Malek et al., 2003), p18^{INK4c}-CDK6 (Jeffrey et al., 2000), CSL-Notch-Mastermind (Wilson and Koval, 2006), Nas6-Rpt3 (Nakamura et al., 2007b), and Gankyrin-S6 (Nakamura et al., 2007a).

Recently, Plückthun and coworkers have chosen the traits of ankyrin repeat proteins for consensus-based protein library design (Binz et al., 2003). Heeding the lessons of nature, they created DNA libraries that encode DARPins consisting of different numbers of repeat modules and special capping repeats attached to the N and C termini of the protein (Figure 1B). Therefore, the molecules are termed Nx_C, where x indicates the number of ankyrin repeat modules. The consensus-designed idealized repeat module consists of conserved residues essential for the formation and stabilization of the repeat module itself and the interrepeat stacking interactions. These positions are invariant and build the framework of the individual repeats. In addition, positions not contributing to the structural integrity of the basic fold were defined to be adaptive surface residues (Figure 1A). These residues are positioned at the typical interface region of the natural ankyrin repeat, namely the tip of the β turn and along the exposed surface of the first α -helix (Figure 1C). At these positions, the introduced diversification allows any amino acid except proline and glycine (structurally unfavorable) and cysteine (could form unwanted disulfide bonds). This design resulted in a virtually unlimited repertoire of molecular surfaces.

Selection Process

Rapid and easy production is a prerequisite for alternative scaffolds suitable for cocrystallization. Repeat protein libraries (Binz et al., 2004; Binz et al., 2003) can be selected using powerful library selection technologies such as ribosome display (RD) (Hanes and Plückthun, 1997) or phage display (Smith, 1985), finally yielding specific binders to the target protein. RD is a complete in vitro technology based on in vitro translation in which noncovalent ternary complexes consisting of mRNA, ribosome, and nascent protein chain are formed. These ternary complexes can then be tested for binding to a particular target protein. This system thereby guarantees the coupling of the genotype (mRNA) to the phenotype (protein) and allows immediate access to the genetic information of the binders. To achieve this, the coding sequence of the protein is genetically fused to a C-terminal spacer, allowing the correct folding of the nascent polypeptide chain, and the construct lacks a Stop codon, thus preventing release of the mRNA and the polypeptide from the ribosome. Low temperature and high concentrations of magnesium further stabilize the ternary complexes. The disassembly of the complexes results from depletion of magnesium, and the DNA is recovered via reverse transcription. RD is particularly well suited for large libraries, because no transformation steps limit the applicable library size. Furthermore, as a cell-free system, it enables the selection of cytotoxic proteins or of proteins with limited in vivo stability.

All the DARPins presented here were isolated from an N3C library and were selected by RD. DARPins could also be successfully selected from an N2C library, but these did not yield crystals. Ankyrin repeat proteins consisting of four to five repeats in total are very abundant in nature (Bork, 1993) and also many of the solved natural ankyrin repeat structures in complex with their target display such repeat numbers. The target proteins chosen here were all recombinantly expressed, purified, and either in vivo-biotinylated at a specific lysine in the AviTag sequence (MBP, APH, and Caspase-2), in vitro-biotinylated at diverse lysines on the surface (AcrB), or GST-tagged (Plk-1) for immobilization. Within a few RD cycles (usually 3 to 4), it was possible to enrich the resulting residual library with proteins that specifically and with high affinity bind to the target.

A subset of the DARPins obtained by RD was then tested individually for binding using approaches suitable for the particular target protein or the intended use of the binder. For MBP (Binz et al., 2004) and Plk-1 (Bandeiras et al., 2008), a standard ELISA was applied to identify a high-affinity binder. In the case of caspase-2, an in vitro enzymatic activity test combined with a standard ELISA yielded an inhibitor (Schweizer et al., 2007). For APH (Amstutz et al., 2005; Kohl et al., 2005) and AcrB (Sennhauser et al., 2007), an in vivo assay was applied in which the inhibition of the target protein through the DARPins resulted in hypersusceptibility of the bacteria. Normally, the DNA of around 20 DARPins obtained from these assays was sequenced per target. The pools often seemed already quite enriched, revealing groups with similar sequences, although identical sequences were rarely observed. Depending on the sequence variability, five to ten complexes were then purified (yields between 50 and 200 mg/l), characterized, and subjected to crystallization.

Structures of DARPins in Complex with Different Antigens

Maltose-Binding Protein: Proof of Principle

As a model protein for the approach of selecting a specific DARPins protein by RD, the maltose binding protein from *Escherichia coli* was chosen and represents the first successful example for the selection and crystallization of a particular target protein in complex with a DARPins. The structure of the complex was determined at 2.3 Å resolution and the phases for this structure could be determined using a known structure of an unselected DARPins previously solved (Kohl et al., 2003). This was possible because the size of the DARPins (~18 kDa) relative to the size of maltose-binding protein (MBP) (~43 kDa) was appropriate, and there was only one complex in the asymmetric unit. Off7 binds the open form of MBP not involving the sugar-binding cleft but binding three helices at one side of the elongated MBP (Figure 2A). Four lysines forming a positive surface patch which account for 60% of the buried surface area of MBP upon complexation interact closely with off7 (Figure 3A).

This first structure showed that the interface between the selected DARPins off7 and MBP indeed involves primarily residues from all three randomized repeat modules, with the randomized repeat module 2 contributing most to the binding (Figure 4A). The two capping repeats are not involved in the interaction. The binding interface is comparable to the lower limit of natural protein-protein interactions (Jones and Thornton, 1996) (Table 1) and is formed by the concave surface of off7 and a convex surface of

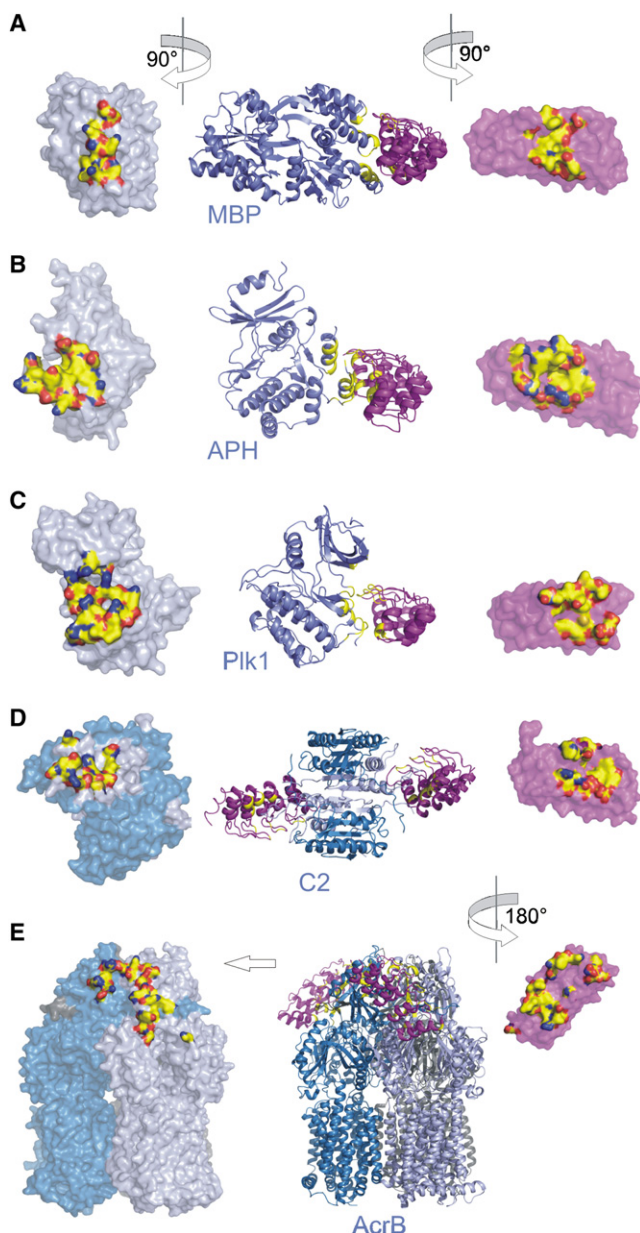


Figure 2. Open Book Illustrations of the Interaction Surfaces of DARPins and their Respective Target Molecules

(A–D) Center: ribbon diagrams of the X-ray structures of the individual complexes. Target proteins and DARPins are colored blue and purple, respectively. Interfaces are highlighted in yellow. Left: the targets have been rotated clockwise. Right: the DARPins have been rotated counterclockwise. Residues that approach within 4 Å of the interaction partner are colored by element (C, N, and O atoms are colored yellow, blue, and red, respectively). (A) DARPin off7 in complex with MBP (PDB entry code 1svx). (B) DARPin 3a in complex with APH (PDB entry code 2bkk). (C) DARPin 3H10 in complex with Plk-1 (PDB entry code 2v5q). (D) DARPin F8 in complex with caspase-2 (PDB entry code 2p2c). Caspase-2 is colored dark blue for the α subunits and light blue for the β subunits. (E) DARPin 110819 in complex with AcrB (PDB entry code 2j8s). The AcrB subunits A, B, and C are colored dark blue, light blue, and gray, respectively. Left: AcrB has been translated. Right: the DARPin has been rotated counterclockwise. Residues that approach within 4 Å of the DARPin are colored by element.

MBP. It involves mainly amino acids at the randomized positions. Off7 contains three framework mutations, of which one (H125Y) is involved in MBP binding. The interface on the DARPin is further characterized by seven aromatic residues, four of which are tyrosines, which account for 72% of the buried surface area of off7 upon complexation.

Aminoglycoside Phosphotransferase (3')-IIIa: The First Inhibiting DARPin

The bacterial kinase aminoglycoside phosphotransferase (3')-IIIa (APH) mediates antibiotic resistance to many pathogenic bacteria and is structurally homologous to eukaryotic protein kinases. A highly affine and potent intracellular kinase inhibitory DARPin, named 3a, was selected (Amstutz et al., 2005), and the structure of APH/3a with bound ADP was determined at a resolution of 2.15 Å with two heterodimeric complexes in the asymmetric unit. The DARPin binds to two helical segments in the C-terminal lobe of APH (Figures 2B and 3B).

On the DARPin surface, the interaction residues are located mainly on the N-terminal capping repeat, on repeat modules 1 and 2, and, to a lesser extent, on module 3 (Figure 4A). As for MBP/off7, there are no interactions involving the C-terminal capping repeat, and as a consequence there was only weak density visible for this part of the molecule, indicating flexibility.

Whereas the structure of the DARPin is not affected by the binding (Figure 4B), 3a binds to the two helices A and B in the α -helical C-terminal lobe of APH, thereby stabilizing helix D in a conformation unable to bind substrate (Figure 5A). In addition, the aminoglycoside positioning loop connecting helices A and B is disordered. Therefore, the structure of the C-terminal lobe of the complex differs significantly from the structure of APH alone. This affects in particular the C-terminal amino acid Phe264 of helix D which is critically involved in substrate binding as revealed by the kanamycin A bound APH structure (Hon et al., 1997). In the DARPin complex, this interaction is distorted, and thus the enzyme is unable to bind the antibiotic.

Polo-Like Kinase-1: An Arduously Crystallizable Target

The Ser/Thr protein kinase polo-like kinase-1 (Plk-1) is a key regulator of mitosis and is a well-validated drug target in cancer therapy. Crystallization attempts for various constructs of the wild-type kinase domain of Plk-1 using a multiparallel cloning and expression approach have failed in the past. Therefore, binders from a DARPin library selective for Plk-1 have been generated and have been used in cocrystallization trials. In this way, the first structure of wild-type apo Plk-1 could be determined in complex with the DARPin 3H10 at 2.3 Å resolution with two complexes in the asymmetric unit. The DARPin binds to the C-terminal lobe of the kinase directly below the ATP binding site (Figure 2C). The binding epitope recognized by DARPin 3H10 is especially rich in Arg, Lys, and Glu residues (66% of the buried surface area) enabling crystallization by masking a surface patch of Plk-1 that is unfavorable for forming crystal contacts similar to the surface residue mutation method (Figure 3C). Such residues have high conformational freedom and are considered counterproductive for crystal contact formation due to a loss of entropy (Derewenda and Vekilov, 2006). As revealed by isothermal calorimetry measurements, the overall interaction is endothermic and apparently driven by entropy alone.

The three repeat modules are solely responsible for the interaction, and again there are mainly residues at the randomized

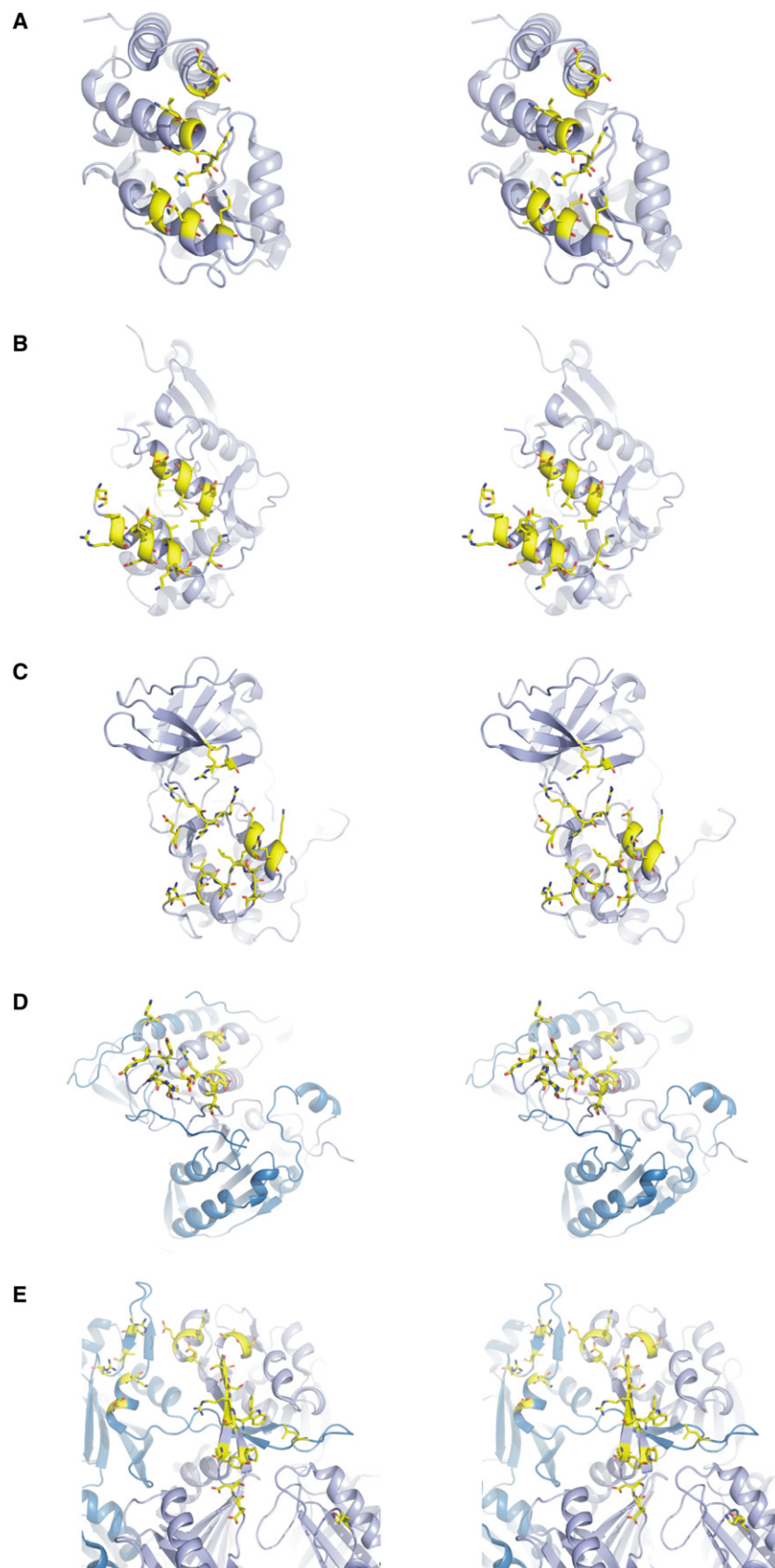


Figure 3. Comparison of the Epitopes on the Different Target Molecules

Stereo view of the targets of the complexes. The molecules are shown in ribbon representation in the same orientation as in the left panel of Figure 2. Interfacing residues (4 Å distance cutoff) are shown as sticks and are colored by element.

(A) MBP.

(B) APh.

(C) Plk-1.

(D) Caspase-2.

(E) Close-up view of AcrB.

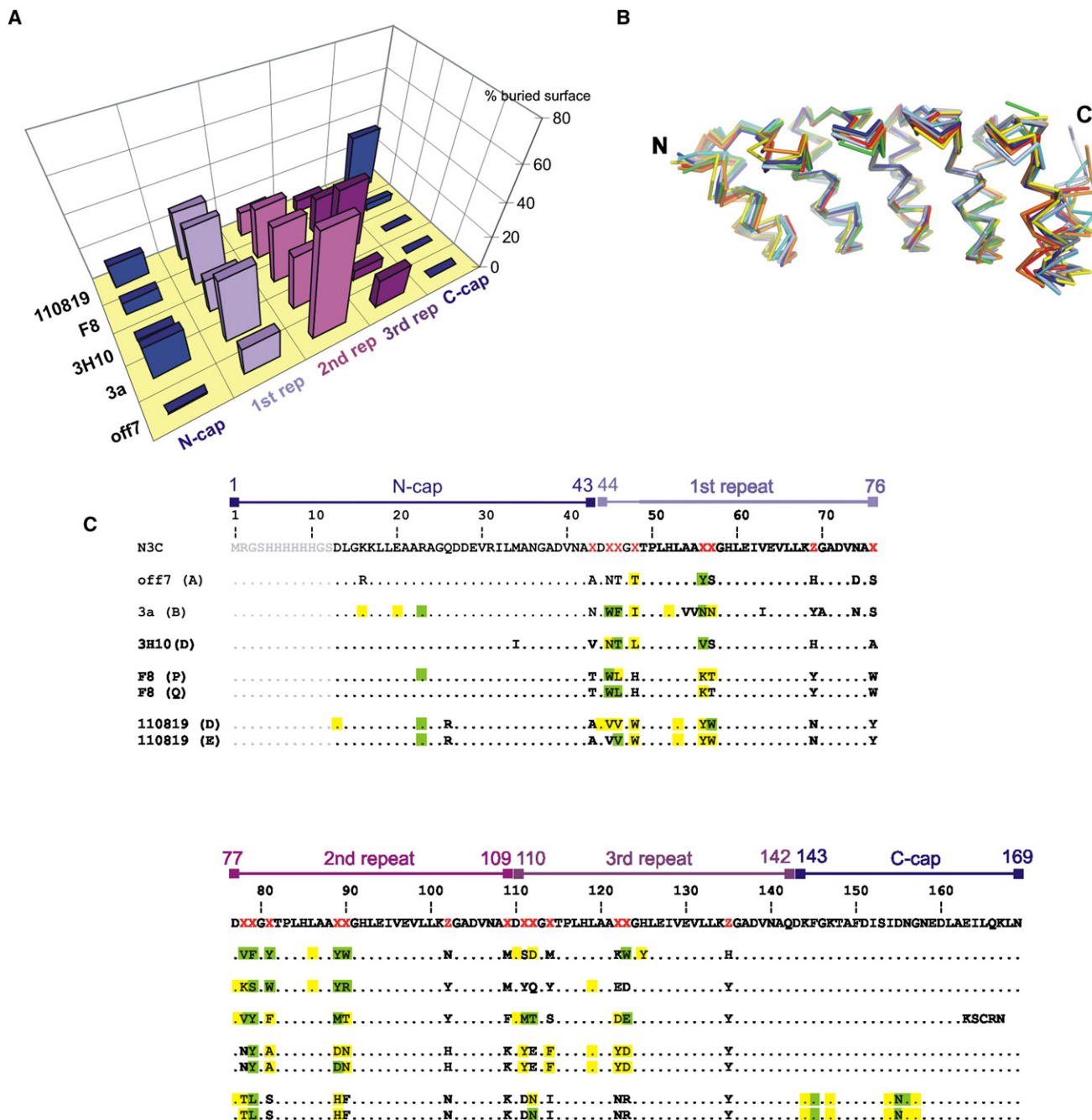


Figure 4. Structural Differences between the Binding DARPins

(A) Contribution of each repeat to the buried surface area of the paratope.

(B) Structures of the DARPins bound to the diverse target molecules. The C α atoms of the DARPins were least-squares superimposed using the program coot and are shown in worm representation. The termini are labeled. The DARPins are colored as follows: red, unselected DARPin; orange, off7; yellow, 3a; green, 3H10; cyan and light blue, F8 molecules P and Q; dark blue and gray, 110819 molecules D and E.

(C) Amino acid sequence alignment of the DARPins. Positions marked with X and Z in the consensus represent the randomized positions where in the case of X any amino acid (except Cys, Gly, and Pro) or in the case of Z Asn, His, and Tyr is allowed. Residues marked in yellow and green indicate those involved in binding of the target in the crystal structure (distance < 4 Å; in green: accounts for more than 5% of the total buried surface area upon complex formation).

positions involved (Figure 4A). The N-terminal capping repeat is not involved in binding. Interestingly, the C-terminal capping repeat is fully disordered in the structure either due to the binding of the kinase or due to clashing with the noncrystallographic symmetry-related DARPin during crystal formation.

Caspase-2: The First Specific Caspase-2 Inhibitor

Caspases are key molecules in the signaling of apoptosis and inflammation. The DARPin F8 was found to inhibit caspase-2 with a subnanomolar inhibition constant. In addition, testing inhibition against other caspases revealed that F8 was highly specific

Table 1. Structural Features of the Interfaces of DARPin-Protein Complexes

Complex	K_D (nM) ^b	Resolution (Å)	N_{res}			Buried Surface Area (Å ²)			Percentage of Total Surface Area (%)			
			Target	DARPin	Total	Target	DARPin	Total	Target	DARPin	N_{HB}	N_{SB}
1svx												
MBP-off7 (B-A) ^a	4.4	2.3	19	21	40	649	615	1264	4.2	8.1	8	2
2bkk												
APH-3a (A-B)	1.7	2.15	21	27	48	925	850	1775	6.7	10.8	15	8
2v5q												
Plk1-3H10 (A-D)	75	2.3	21	25	46	857	885	1742	6.0	14.3	14	9
2p2c												
C2-F8 (BDA-P)	4.1	3.24	25	26	51	937	885	1822	12.1	10.2	8	9
C2-F8 (DBC-Q)			23	25	48	830	809	1639	10.5	9.7	6	9
2j8s												
AcrB-110819 (BA-D)	28	2.54	35	33	68	1142	1186	2328	2.5	15.3	12	3
AcrB-110819 (AC-E)			32	31	63	969	1017	1986	2.2	13.4	11	2
Antibody-antigen ^c						777 ± 135					8.2 ± 0.7	
Heterodimeric protein-protein complexes ^c						983 ± 582					11.1 ± 2.7	

Analyzed with the Protein Interfaces, Surfaces and Assemblies service PISA. The numbers may be at variance with those found by the original authors owing to different programs used. Only the interface features of one complex are shown when there were several complexes in the asymmetric unit. N_{res} – number of residues from the specified molecule that take part in the interface; N_{HB} – number of potential hydrogen bonds across the interface; N_{SB} – number of potential salt bridges across the interface.

^a Characters in parentheses show the chains participating in the interface as designated in the PDB file (target-DARPin).

^b Affinities were determined by BLAcore technology. For Plk1-3H10 a K_D measured by competition ELISA was described.

^c According to Jones and Thornton (1996). The values for N_{HB} were calculated from the tabulated hydrogen bonds/100 Å² ΔASA.

exclusively for caspase-2. The structure of the caspase-2/F8 complex at 3.24 Å resolution revealed the molecular basis for the specificity as well as the allosteric mechanism of inhibition which was also verified by kinetic analysis. The asymmetric unit contained three caspase-2 dimers with two bound DARPins each (Figure 2D). The main contact surface of caspase-2 is formed by the 381 loop, involving diverse residues (Figure 3D). Minor interactions involve the N-terminus of the other β-subunit of the caspase. Regarding F8, all three repeat modules and one framework residue in the N-terminal capping repeat are forming the interface on F8 (Figure 4A).

The 381 loop forms one side of the active site cleft (Figure 5B). The conformation of this loop is slightly shifted when compared

with the caspase-2/peptide inhibitor structure (Schweizer et al., 2003), resulting in an opening of the active site cleft. Additional rearrangements in the surrounding of the active site cleft and the side chain of Tyr79 of F8 occupying partly the substrate sub-pocket S5 of caspase-2 are the main reasons for the inhibitory property of F8. The caspase-2/F8 complex shows that the allosteric inhibition of caspase-2 is extremely specific, which has significant advantages over orthosteric ligands which often lack specificity.

AcrB: The First Integral Membrane Protein in Complex with a DARPin

The multidrug exporter AcrB, the inner membrane component of the AcrAB-TolC drug efflux system in *Escherichia coli* is the

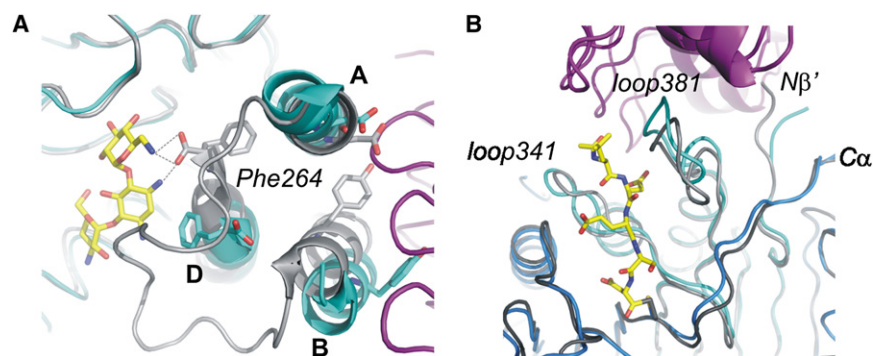


Figure 5. Conformations of the Target Molecules

(A) Close-up view of the superposition of the DARPin binding region of APH in complex with DARPin inhibitor 3a (blue and purple) and wtAPH (gray). The APH/3a complex displays a distorted drug binding site. The α helices A and B in the C-terminal lobe of APH interacting with the DARPin are shifted, and the aminoglycoside positioning loop between these two helices is disordered. This rearrangement leads to a 5- to 7-Å shift of helix D with respect to wtAPH, which includes the C-terminal residue Phe264. Kanamycin A bound to wtAPH is shown in yellow, and hydrogen bonds to the mainchain carboxyl group of Phe264 are indicated as dashed lines.

(B) Close-up view of the active-site cleft region of caspase-2. Superimposed caspase-2 bound to Ac-LDESD-CHO (colored in gray and yellow) and to the DARPin F8 (colored in blue and red) is shown. In the DARPin-bound form, the loop 381 and the C-terminal part of the α subunit (Cα), which includes the active site cysteine, are shifted with respect to the peptide bound form.

major system contributing to efflux-caused drug resistance in this organism. This protein actively detoxifies the intracellular space by exporting drugs to the cell exterior while importing protons. To explain the molecular mechanism of this process, the availability of high-resolution structural information is a prerequisite. For this, and to show the applicability of the DARPIn combinatorial libraries and of the ribosome display selection method to membrane proteins, AcrB was chosen as model system. Although AcrB can be crystallized on its own due to its large periplasmic domain protruding from the lipid bilayer, the crystals of the AcrB/110819 complex diffracted to 2.54 Å resolution, substantially better than all the previously published crystals of AcrB alone (Murakami et al., 2002). In the structure, the DARPIn binds to the periplasmic part of AcrB and due to the distinctly different conformations of the subunits only two DARPins bind to the trimeric protein (Figure 2E). The epitope on AcrB is mainly formed by a β sheet connecting the pore domain with the TolC docking domain, which are subdomains of the periplasmic portion of the protein (Figure 3E). An additional interaction is formed involving a helical segment and a β strand of the adjacent preceding subunit of AcrB via hydrophobic interactions and two hydrogen bonds.

All five repeats (including both capping repeats) of the DARPIn are involved in the interaction (Figure 4A). Comprising around 1100 Å² buried surface area on both the DARPIn and AcrB, and 68 residues in total, this complex showed the largest buried surface area for a DARPIn complex so far (Table 1).

Interestingly, the structure revealed a new conformation of the molecule where the three subunits of AcrB were locked in different conformations, revealing distinct channels in each subunit. Furthermore, the structure indicated how in the protein the coupling between the channel access, exit, and the putative proton translocation site is achieved. It suggests a transport pathway through the identified channels in the individual subunits by a rotary mechanism. At the same time, this new conformation was also observed without a bound DARPIn by two individual groups (Murakami et al., 2006; Seeger et al., 2006). In further studies, DARPIn 110819 also allowed the crystallization of otherwise uncrystallizable mutants (K.M. Pos, unpublished data).

General Properties of DARPIn-Target Interfaces

The structural properties forming the surface of the target proteins recognized by DARPins vary from complex to complex (Figure 2). A common trait between the different interfaces, however, is the overall curvature of the target protein interface area enclosed by DARPins. This curvature is complementary to the curvature of the randomized surface area of the DARPIn. Furthermore, all three internal randomized repeat modules are involved in the interaction in all five complexes (Figure 4A). As set out in the design, the predominant interaction is mediated by residues in the randomized positions (Figure 4C); this is the reason for the high specificity of the selected molecules. Interestingly, the last residue of the individual repeat module, namely randomized position 33, is not involved in any interface. In contrast, position 3 is heavily buried in most of the repeat modules. In all complexes, framework residues also participate in the interaction with the target protein. In the case of the MBP and AcrB binders, there is a mutated framework position H125Y and D155N, respectively, contributing to the interaction. Such frame-

work mutations can be explained as a result of the numerous PCR cycles the DARPIn-encoding DNA passes during the selection process.

The involvement of the capping repeats in the interface differs greatly between the different complexes. Whereas the N-terminal capping repeat is well ordered in all structures and involved in binding in the APH, caspase-2, and AcrB complexes, the C-terminal capping repeat is fully disordered in the Plk-1 complex and partially disordered in the APH and caspase-2 complex. It is solely involved in the interface of the AcrB/110819 complex, where 7 residues of the C-terminal capping repeat are participating in the interface. In having the first interacting residue at position 13 and the last at position 157, the full width of the 169 amino acid residue DARPIn is used in this complex. This provides the most extensive interaction described to date.

The structures discussed here show that DARPins are not restricted to recognize any specific amino acid sequence or structure on the target. In all cases, the DARPins recognize native, nonlinear epitopes (Figure 3). There is no particular chemical character found neither for the epitopes (target interface) nor for the paratopes (DARPIn interface). There are numerous aromatic residues on the paratopes, but they seem to be less preferred than in antibody paratopes (Padlan, 1990). An exception is the MBP-off7 complex where the aromatic residues account for 72% of the buried surface area of the DARPIn. The proportion of aromatic residues in the epitopes is low, a fact that has been observed in antibody-antigen interactions as well (Davies and Cohen, 1996). Furthermore, charged residues seem to be important for the interaction, accounting for nearly half of the contact residues of the epitopes. Whereas all DARPins described here are composed of five repeat modules, the total interface sizes on the target molecules differ widely. The MBP/off7 complex presents by far the smallest interface, comprising 40 residues and a combined buried surface area of only 1264 Å² (Table 1). In contrast, the largest interface observed in the AcrB/110819 complex comprises 63 and 68 residues and buries 1986 and 2328 Å² of the AcrB surface in the two interfaces formed; all other complex interfaces are in between these values. However, the interface sizes show great similarities to naturally occurring ankyrin repeat protein, heterodimeric protein-protein, and antibody-antigen complexes (Table 1).

Structural Rearrangements in the Target Proteins and the DARPins

An often raised argument in the field of chaperone-assisted crystallization is that binders might induce a conformational change in the target molecules. Because the crystal structures of all target proteins chosen are also available without a binding partner, it is possible to investigate conformational changes of the targets on DARPIn binding. The structures of MBP in the open conformation without a ligand (PDB ID code: 1LLS [Rubin et al., 2002]) and complexed MBP are very similar, with an overall C α rmsd below 1 Å. Also, in the case of AcrB, the structures obtained in 2006 (Murakami et al., 2006; Seeger et al., 2006) are remarkably similar, considering the size of the protein (rmsd for 3110 C α atoms $\sim$ 1 Å). In addition, the observed asymmetry in the crystal structure could also be detected in solution (Sennhauser et al., 2007). The overall structure of Plk-1 is also very similar to two structures of a mutant solved at the same time,

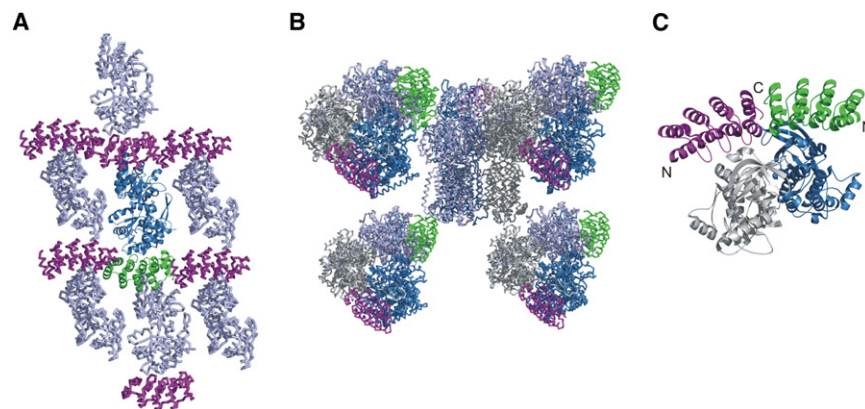


Figure 6. Diverse Crystal Contacts Mediated by DARPins

(A) Crystal packing in the MBP-DARPin complex. The crystals belong to space group $P2_1$. One complex is shown in ribbon representation and colored dark blue (MBP) and green (DARPin). The neighboring molecules are shown in worm representation in light blue and purple. Crystal contacts are mediated mainly between adjacent DARPin and MBP molecules.

(B) Crystal contacts of AcrB molecules and DARPins. The crystals belong to space group $P2_12_12_1$. The AcrB subunits A, B, and C are colored dark blue, light blue, and gray, respectively. One of the two DARPins (in green) bound to the periplasmic part of AcrB mediates essential crystal contacts to the cytoplasmic part of a symmetry-related trimer, resulting in a different packing to AcrB crystallized without a DARPin.

(C) Asymmetric unit content of the Plk-1/DARPin crystal. The two complexes are shown in ribbon representation and are colored gray and blue (Plk-1) and purple and green (DARPin). The N- and C-terminal ends are labeled.

each harboring a different ligand (Kothe et al., 2007), again indicating that there is no binding protein induced conformational change.

The interaction of the DARPin inhibitors with APH and caspase-2 revealed conformations in which the target proteins were inactive. This information is useful for finding novel ways to inhibit the function of an enzyme. The structural rearrangements there are quite considerable. In APH, helices A, B, and D are significantly shifted with respect to kanamycin A-bound APH (Hon et al., 1997), and the aminoglycoside positioning loop connecting helices A and B is disordered (Figure 5A). In caspase-2, the loops surrounding the active site cleft are altered when compared with the caspase-2/peptide inhibitor structure (Schweizer et al., 2003), resulting in an opening of the active site cleft (Figure 5B). Nevertheless, we do not believe that these structural rearrangements are caused by the DARPin, resulting in an induced fit. Recalling the DARPin selection strategy, it is more likely that the DARPins only bind to epitopes as presented by the target proteins in a “lock and key”-like fashion. In this way, the selected DARPins can not only bind to proteins in their predominant conformation but can also trap the proteins in conformations that are probably present in solution but are less populated, thus inhibiting an enzyme in an allosteric way.

In all structures, the differences in the DARPin scaffold are minimal upon binding and showed no essential rearrangements of the global fold compared with an unselected DARPin alone (Kohl et al., 2003) (Figure 4B). In agreement with the participation of the individual repeats in the interaction, the strongest flexibility can be observed in the C-terminal capping repeat. In the DARPin-binding Plk-1, the C-terminal capping repeat cannot even be localized in the crystal structure. The β turn regions of the individual repeat modules also show some minor adaptations when bound to a target protein.

DARPin-Mediated Crystal Contacts

The potential of DARPins to provide crystal contacts is best described for the MBP/off7 and the AcrB/110819 complex. As shown in Figure 6A, in the MBP/off7 complex the lattice interactions involve only DARPin-DARPin and MBP-MBP contacts,

which occur within layers, and the direct binding interface between the two interaction partners is the sole contact form between the layers.

In the crystal formed by the AcrB/110819 complex, crystal contacts are mediated by the DARPin in such a way that, in addition to the interaction with the periplasmic domain of its target, the convex side of the DARPin simultaneously forms crystal contacts to the cytoplasmic polar surface of a symmetry-related molecule of AcrB (Figure 6B). Because the crystal has the space group $P2_12_12_1$, this contact critically contributes to the stability of the lattice (the mentioned contact occurs in all three dimensions) and thus to the higher diffraction limits of the complex crystal over the crystal formed by AcrB alone.

There are some proteins that could be crystallized only when complexed to a cognate Fab fragment (Laver et al., 1990). A similar case is Plk-1 that alone in its apo-form—a structure most useful for drug design—did not crystallize. In complex with the DARPin 3H10, however, Plk-1 did crystallize. Surprisingly, the DARPin makes direct crystal contacts to the other DARPin in the asymmetric unit via the third repeat module as the C-terminal capping repeat is fully disordered (Figure 6C).

It is worth mentioning that cocrystallization with a DARPin clearly was advantageous in the cases of AcrB and Plk-1, whereas the opposite was true for caspase-2. Here, the diffraction quality of the DARPin complex was clearly worse (3.24 Å) than caspase-2 in complex with a peptide inhibitor (1.65 Å) (Schweizer et al., 2003). DARPins containing different numbers of internal repeats might be selected to obtain better diffracting crystals.

Conclusions

We have reviewed the recent results using DARPins as crystallization tools to crystallize difficult target proteins to understand the natural and alternative conformations of these proteins. DARPins are a class of designed binding molecules that represent a very promising alternative to antibodies. They retain the affinity and specificity of antibodies but are far more stable, and their simple architecture allows high-yield bacterial production. Their use in biomedical and biotechnological research is

considerably more advantageous over antibodies. High-affinity DARPins to a given target can rapidly be selected from very large libraries within a few weeks. The use of in vitro selection systems such as phage or ribosome display represents an advantage over the traditional production of monoclonal antibodies in test animals, as the binders are selected to a native protein under defined buffer conditions. Therefore, the in vitro display techniques guarantee that surface epitopes of the presented target protein are selected. The use in structural biology is obvious from the structures described here, and not only enables crystallization of soluble and membrane proteins but also reveals conformations of the target proteins that otherwise are not detected when inhibitors are selected. In the cases of DARPin inhibitors of caspase-2 and APH, the crystals of the complex did not diffract better; rather, the diffraction was worse than that of crystals of the target proteins with small molecule compounds bound to the active site. However, in both cases a conformation of the target protein was revealed in which the protein is inactive. This information is useful for finding new ways to modulate target protein function.

To expand the applicability of the technology in protein crystallography, heavy atom sites can be engineered into the capping repeats. Crystals of target proteins with a heavy atom-loaded DARPin is a great help in phasing using the single anomalous dispersion method. This has already been successfully applied using a DARPin-MBP complex crystal (A. Plückthun and M.G.Grütter, unpublished data).

In conclusion, DARPins have the potential to (1) provide crystals of a target protein that is difficult to crystallize, (2) provide better diffracting crystals of the complex when the target protein alone only poorly diffracts, (3) provide initial phases for molecular replacement in cases where the DARPin and the target protein are not too different in size, (4) provide phases by single anomalous dispersion for complexes formed with heavy atom loaded DARPins, and (5) reveal new insight into the function of a target protein by stabilizing a conformation that in solution is less populated.

ACKNOWLEDGMENTS

Support from the Swiss National Science Foundation in the framework of the National Center of Competence in Research Structural Biology is acknowledged. We would like to thank Patrick Amstutz for critical reading of the manuscript.

REFERENCES

- Amit, A.G., Mariuzza, R.A., Phillips, S.E., and Poljak, R.J. (1986). Three-dimensional structure of an antigen-antibody complex at 2.8 Å resolution. *Science* 233, 747–753.
- Amstutz, P., Binz, H.K., Parizek, P., Stumpp, M.T., Kohl, A., Grütter, M.G., Forrer, P., and Plückthun, A. (2005). Intracellular kinase inhibitors selected from combinatorial libraries of designed ankyrin repeat proteins. *J. Biol. Chem.* 280, 24715–24722.
- Anderson, J., Ptashne, M., and Harrison, S.C. (1984). Cocrystals of the DNA-binding domain of phage 434 repressor and a synthetic phage 434 operator. *Proc. Natl. Acad. Sci. USA* 81, 1307–1311.
- Bandeiras, T.M., Hillig, R.C., Matias, P.M., Eberspaecher, U., Fanghanel, J., Thomaz, M., Miranda, S., Crusius, K., Putter, V., Amstutz, P., et al. (2008). Structure of wild-type Plk-1 kinase domain in complex with a selective DARPin. *Acta Crystallogr. D Biol. Crystallogr.* 64, 339–353.
- Batchelor, A.H., Piper, D.E., de la Brousse, F.C., McKnight, S.L., and Wolberger, C. (1998). The structure of GABPalpha/beta: an ETS domain-ankyrin repeat heterodimer bound to DNA. *Science* 279, 1037–1041.
- Bentley, G.A. (1996). The crystal structures of complexes formed between lysozyme and antibody fragments. *EXS* 75, 301–319.
- Binz, H.K., Stumpp, M.T., Forrer, P., Amstutz, P., and Plückthun, A. (2003). Designing repeat proteins: well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins. *J. Mol. Biol.* 332, 489–503.
- Binz, H.K., Amstutz, P., Kohl, A., Stumpp, M.T., Briand, C., Forrer, P., Grütter, M.G., and Plückthun, A. (2004). High-affinity binders selected from designed ankyrin repeat protein libraries. *Nat. Biotechnol.* 22, 575–582.
- Bork, P. (1993). Hundreds of ankyrin-like repeats in functionally diverse proteins: mobile modules that cross phyla horizontally? *Proteins* 17, 363–374.
- Brotherton, D.H., Dhanaraj, V., Wick, S., Brizuela, L., Domaille, P.J., Volyanik, E., Xu, X., Parisini, E., Smith, B.O., Archer, S.J., et al. (1998). Crystal structure of the complex of the cyclin D-dependent kinase Cdk6 bound to the cell-cycle inhibitor p19INK4d. *Nature* 395, 244–250.
- Davies, D.R., and Cohen, G.H. (1996). Interactions of protein antigens with antibodies. *Proc. Natl. Acad. Sci. USA* 93, 7–12.
- DeLano, W.L. (2002). The PyMOL Molecular Graphics System (San Carlos, CA: DeLano Scientific).
- Derewenda, Z.S. (2004). The use of recombinant methods and molecular engineering in protein crystallization. *Methods* 34, 354–363.
- Derewenda, Z.S., and Vekilov, P.G. (2006). Entropy and surface engineering in protein crystallization. *Acta Crystallogr. D Biol. Crystallogr.* 62, 116–124.
- Desmyter, A., Transue, T.R., Ghahroudi, M.A., Thi, M.H., Poortmans, F., Hamers, R., Muyldermans, S., and Wyns, L. (1996). Crystal structure of a camel single-domain VH antibody fragment in complex with lysozyme. *Nat. Struct. Biol.* 3, 803–811.
- Dutzler, R., Campbell, E.B., and MacKinnon, R. (2003). Gating the selectivity filter in CIC chloride channels. *Science* 300, 108–112.
- Fellouse, F.A., Wiesmann, C., and Sidhu, S.S. (2004). Synthetic antibodies from a four-amino-acid code: a dominant role for tyrosine in antigen recognition. *Proc. Natl. Acad. Sci. USA* 101, 12467–12472.
- Fellouse, F.A., Li, B., Compaan, D.M., Peden, A.A., Hymowitz, S.G., and Sidhu, S.S. (2005). Molecular recognition by a binary code. *J. Mol. Biol.* 348, 1153–1162.
- Fellouse, F.A., Esaki, K., Birtalan, S., Raptis, D., Cancasci, V.J., Koide, A., Jhurani, P., Vasser, M., Wiesmann, C., Kossiakoff, A.A., et al. (2007). High-throughput generation of synthetic antibodies from highly functional minimalist phage-displayed libraries. *J. Mol. Biol.* 373, 924–940.
- Gill, D.S., and Damle, N.K. (2006). Biopharmaceutical drug discovery using novel protein scaffolds. *Curr. Opin. Biotechnol.* 17, 653–658.
- Hanes, J., and Plückthun, A. (1997). In vitro selection and evolution of functional proteins by using ribosome display. *Proc. Natl. Acad. Sci. USA* 94, 4937–4942.
- Hogbom, M., Eklund, M., Nygren, P.A., and Nordlund, P. (2003). Structural basis for recognition by an in vitro evolved affibody. *Proc. Natl. Acad. Sci. USA* 100, 3191–3196.
- Hon, W.C., McKay, G.A., Thompson, P.R., Sweet, R.M., Yang, D.S., Wright, G.D., and Berghuis, A.M. (1997). Structure of an enzyme required for aminoglycoside antibiotic resistance reveals homology to eukaryotic protein kinases. *Cell* 89, 887–895.
- Hunte, C., and Michel, H. (2002). Crystallisation of membrane proteins mediated by antibody fragments. *Curr. Opin. Struct. Biol.* 12, 503–508.
- Hunte, C., Koepke, J., Lange, C., Rossmann, T., and Michel, H. (2000). Structure at 2.3 Å resolution of the cytochrome bc(1) complex from the yeast *Saccharomyces cerevisiae* co-crystallized with an antibody Fv fragment. *Structure* 8, 669–684.

- Huxford, T., Huang, D.B., Malek, S., and Ghosh, G. (1998). The crystal structure of the IkappaBalpha/NF-kappaB complex reveals mechanisms of NF-kappaB inactivation. *Cell* 95, 759–770.
- Iwata, S., Ostermeier, C., Ludwig, B., and Michel, H. (1995). Structure at 2.8 Å resolution of cytochrome c oxidase from *Paracoccus denitrificans*. *Nature* 376, 660–669.
- Jacobs, M.D., and Harrison, S.C. (1998). Structure of an IkappaBalpha/NF-kappaB complex. *Cell* 95, 749–758.
- Jeffrey, P.D., Tong, L., and Pavletich, N.P. (2000). Structural basis of inhibition of CDK-cyclin complexes by INK4 inhibitors. *Genes Dev.* 14, 3115–3125.
- Jones, S., and Thornton, J.M. (1996). Principles of protein-protein interactions. *Proc. Natl. Acad. Sci. USA* 93, 13–20.
- Kohl, A., Binz, H.K., Forrer, P., Stumpp, M.T., Plückthun, A., and Grütter, M.G. (2003). Designed to be stable: crystal structure of a consensus ankyrin repeat protein. *Proc. Natl. Acad. Sci. USA* 100, 1700–1705.
- Kohl, A., Amstutz, P., Parizek, P., Binz, H.K., Briand, C., Capitani, G., Forrer, P., Plückthun, A., and Grütter, M.G. (2005). Allosteric inhibition of aminoglycoside phosphotransferase by a designed ankyrin repeat protein. *Structure* 13, 1131–1141.
- Koide, A., Gilbreth, R.N., Esaki, K., Tereshko, V., and Koide, S. (2007). High-affinity single-domain binding proteins with a binary-code interface. *Proc. Natl. Acad. Sci. USA* 104, 6632–6637.
- Kortt, A.A., Malby, R.L., Caldwell, J.B., Gruen, L.C., Ivancic, N., Lawrence, M.C., Howlett, G.J., Webster, R.G., Hudson, P.J., and Colman, P.M. (1994). Recombinant anti-sialidase single-chain variable fragment antibody. Characterization, formation of dimer and higher-molecular-mass multimers and the solution of the crystal structure of the single-chain variable fragment/sialidase complex. *Eur. J. Biochem.* 227, 151–157.
- Kothe, M., Kohls, D., Low, S., Coli, R., Cheng, A.C., Jacques, S.L., Johnson, T.L., Lewis, C., Loh, C., Nonomiya, J., et al. (2007). Structure of the catalytic domain of human polo-like kinase 1. *Biochemistry* 46, 5960–5971.
- Laver, W.G., Air, G.M., Webster, R.G., and Smith-Gill, S.J. (1990). Epitopes on protein antigens: misconceptions and realities. *Cell* 61, 553–556.
- Lawson, D.M., Artymiuk, P.J., Yewdall, S.J., Smith, J.M., Livingstone, J.C., Treffry, A., Luzzago, A., Levi, S., Arosio, P., Cesareni, G., et al. (1991). Solving the structure of human H ferritin by genetically engineering intermolecular crystal contacts. *Nature* 349, 541–544.
- Malek, S., Huang, D.B., Huxford, T., Ghosh, S., and Ghosh, G. (2003). X-ray crystal structure of an IkappaBbeta x NF-kappaB p65 homodimer complex. *J. Biol. Chem.* 278, 23094–23100.
- Murakami, S., Nakashima, R., Yamashita, E., and Yamaguchi, A. (2002). Crystal structure of bacterial multidrug efflux transporter AcrB. *Nature* 419, 587–593.
- Murakami, S., Nakashima, R., Yamashita, E., Matsumoto, T., and Yamaguchi, A. (2006). Crystal structures of a multidrug transporter reveal a functionally rotating mechanism. *Nature* 443, 173–179.
- Nakamura, Y., Nakano, K., Umehara, T., Kimura, M., Hayashizaki, Y., Tanaka, A., Horikoshi, M., Padmanabhan, B., and Yokoyama, S. (2007a). Structure of the oncoprotein gankyrin in complex with S6 ATPase of the 26S proteasome. *Structure* 15, 179–189.
- Nakamura, Y., Umehara, T., Tanaka, A., Horikoshi, M., Padmanabhan, B., and Yokoyama, S. (2007b). Structural basis for the recognition between the regulatory particles Nas6 and Rpt3 of the yeast 26S proteasome. *Biochem. Biophys. Res. Commun.* 359, 503–509.
- Ostermeier, C., Iwata, S., Ludwig, B., and Michel, H. (1995). Fv fragment-mediated crystallization of the membrane protein bacterial cytochrome c oxidase. *Nat. Struct. Biol.* 2, 842–846.
- Padlan, E.A. (1990). On the nature of antibody combining sites: unusual structural features that may confer on these sites an enhanced capacity for binding ligands. *Proteins* 7, 112–124.
- Rothe, C., Urlinger, S., Lohning, C., Prassler, J., Stark, Y., Jager, U., Hubner, B., Bardroff, M., Pradel, I., Boss, M., et al. (2008). The human combinatorial antibody library HuCAL GOLD combines diversification of all six CDRs according to the natural immune system with a novel display method for efficient selection of high-affinity antibodies. *J. Mol. Biol.* 376, 1182–1200.
- Rubin, S.M., Lee, S.Y., Ruiz, E.J., Pines, A., and Wemmer, D.E. (2002). Detection and characterization of xenon-binding sites in proteins by ¹²⁹Xe NMR spectroscopy. *J. Mol. Biol.* 322, 425–440.
- Russo, A.A., Tong, L., Lee, J.O., Jeffrey, P.D., and Pavletich, N.P. (1998). Structural basis for inhibition of the cyclin-dependent kinase Cdk6 by the tumour suppressor p16INK4a. *Nature* 395, 237–243.
- Schweizer, A., Briand, C., and Grütter, M.G. (2003). Crystal structure of caspase-2, apical initiator of the intrinsic apoptotic pathway. *J. Biol. Chem.* 278, 42441–42447.
- Schweizer, A., Roschitzki-Voser, H., Amstutz, P., Briand, C., Gulotti-Georgieva, M., Prenosil, E., Binz, H.K., Capitani, G., Baici, A., Plückthun, A., and Grütter, M.G. (2007). Inhibition of caspase-2 by a designed ankyrin repeat protein: specificity, structure, and inhibition mechanism. *Structure* 15, 625–636.
- Sedgwick, S.G., and Smerdon, S.J. (1999). The ankyrin repeat: a diversity of interactions on a common structural framework. *Trends Biochem. Sci.* 24, 311–316.
- Seeger, M.A., Schiefner, A., Eicher, T., Verrey, F., Diederichs, K., and Pos, K.M. (2006). Structural asymmetry of AcrB trimer suggests a peristaltic pump mechanism. *Science* 313, 1295–1298.
- Sennhauser, G., Amstutz, P., Briand, C., Storchenegger, O., and Grütter, M.G. (2007). Drug export pathway of multidrug exporter AcrB revealed by DARPIn inhibitors. *PLoS Biol.* 5, e7.
- Smith, G.P. (1985). Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* 228, 1315–1317.
- Stumpp, M.T., and Amstutz, P. (2007). DARPins: a true alternative to antibodies. *Curr. Opin. Drug Discov. Dev.* 10, 153–159.
- Wahlberg, E., Lendel, C., Helgstrand, M., Allard, P., Dincbas-Renqvist, V., Hedqvist, A., Berglund, H., Nygren, P.A., and Hard, T. (2003). An affibody in complex with a target protein: structure and coupled folding. *Proc. Natl. Acad. Sci. USA* 100, 3185–3190.
- Wilson, J.J., and Kovall, R.A. (2006). Crystal structure of the CSL-Notch-Mastermind ternary complex bound to DNA. *Cell* 124, 985–996.
- Ye, J.D., Tereshko, V., Frederiksen, J.K., Koide, A., Fellouse, F.A., Sidhu, S.S., Koide, S., Kossiakoff, A.A., and Piccirilli, J.A. (2008). Synthetic antibodies for specific recognition and crystallization of structured RNA. *Proc. Natl. Acad. Sci. USA* 105, 82–87.
- Zhou, Y., Morais-Cabral, J.H., Kaufman, A., and MacKinnon, R. (2001). Chemistry of ion coordination and hydration revealed by a K⁺ channel-Fab complex at 2.0 Å resolution. *Nature* 414, 43–48.

High-affinity binders selected from designed ankyrin repeat protein libraries

H Kaspar Binz^{1,2}, Patrick Amstutz^{1,2}, Andreas Kohl^{1,2}, Michael T Stumpp¹, Christophe Briand¹, Patrik Forrer¹, Markus G Grütter¹ & Andreas Plückthun¹

We report here the evolution of ankyrin repeat (AR) proteins *in vitro* for specific, high-affinity target binding. Using a consensus design strategy, we generated combinatorial libraries of AR proteins of varying repeat numbers with diversified binding surfaces. Libraries of two and three repeats, flanked by 'capping repeats,' were used in ribosome-display selections against maltose binding protein (MBP) and two eukaryotic kinases. We rapidly enriched target-specific binders with affinities in the low nanomolar range and determined the crystal structure of one of the selected AR proteins in complex with MBP at 2.3 Å resolution. The interaction relies on the randomized positions of the designed AR protein and is comparable to natural, heterodimeric protein-protein interactions. Thus, our AR protein libraries are valuable sources for binding molecules and, because of the very favorable biophysical properties of the designed AR proteins, an attractive alternative to antibody libraries.

Repeat proteins are ubiquitous binding molecules fundamental to many biological processes^{1–3}. Their modular architecture is presumably the key to their evolutionary success⁴. Repeat proteins are characterized by consecutive homologous structural units (repeats), which stack to form an elongated protein domain with a continuous hydrophobic core³. In principle, this architecture allows their binding specificities to evolve not only by point mutations but also by insertion, deletion or shuffling of repeats⁵. This evolutionary strategy might enable repeat proteins to acquire new functions by adjusting their surface without jeopardizing their overall topology. AR proteins are one prominent repeat protein family illustrating the binding versatility of repeat proteins. They occur throughout all phyla and mediate protein-protein interactions in the nucleus or cytoplasm, or while anchored to the membrane or when secreted into the extracellular space⁶. AR proteins are built from stacked, 33 amino acid repeats, each forming a β -turn that is followed by two antiparallel α -helices and a loop reaching the β -turn of the next repeat⁷. In most known complexes, the β -turn and the first α -helix mediate the interactions with the target, and different numbers of adjacent repeats are involved in binding⁷. The reported target binding affinities of natural AR proteins are in the low nanomolar range^{8,9}.

In biotechnology and biomedical research, antibodies and fragments thereof are the most widely used specific, high-affinity binding molecules. Antibodies can be generated against essentially any target either by immunization or by using natural or rationally designed antibody libraries *in vitro*^{10,11}. Yet, many antibodies have relatively low expression yields, a tendency to aggregate and a dependence on disulfide bonds for stability. An ideal alternative protein scaffold would have none of these drawbacks while still exhibiting the same affinity and specificity

as antibodies. Previous attempts to generate alternative binding molecules relied on either loop or surface randomization of protein scaffolds, which are typically small and always fixed in dimension^{12–14}.

We present here the results of a different strategy¹⁵ to design and select alternative protein scaffolds, relying on the modularity of AR proteins (Fig. 1). We generated combinatorial libraries of consensus-designed AR proteins of varying sizes (that is, varying repeat numbers) with randomized potential interaction surfaces. Unselected library members are very well expressed, soluble, thermodynamically stable and show the typical AR domain fold^{16,17}. Here, we show the successful selection of binding molecules from these libraries, proving that our design strategy works. We selected specific binders with high affinities for the *Escherichia coli* maltose binding protein (MBP) and two eukaryotic mitogen-activated protein kinases (MAPKs). The crystal structure of one of the selected binders in complex with MBP was determined, revealing atomic level insights into the mode of target interaction of this class of designed binding molecules.

RESULTS

Designed AR protein libraries

We designed a consensus AR module consisting of six diversified potential interaction residues (which can be any amino acid except cysteine, glycine and proline) and 27 framework residues (26 are fixed and one is allowed to be asparagine, histidine or tyrosine). This module was designed from sequence alignments and structural analyses (Fig. 1)¹⁶. The randomized potential interaction residues are located in the β -turn and the first α -helix of the AR module. We cloned varying numbers of this repeat module between capping repeats, which are special terminal repeats of AR domains shielding the hydrophobic

¹Biochemisches Institut, Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland. ²These authors contributed equally to this work. Correspondence should be addressed to A.P. (plueckthun@bioc.unizh.ch) or M.G.G. (gruetter@bioc.unizh.ch).

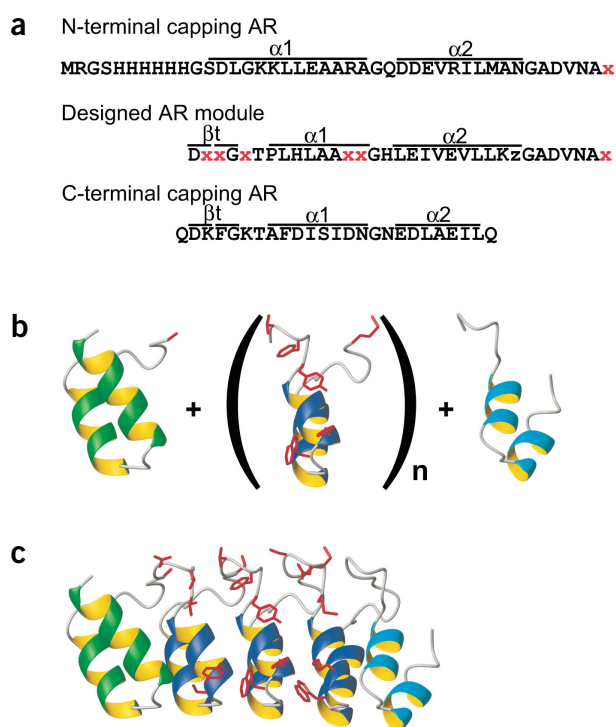


Figure 1 Construction of designed AR protein libraries. **(a)** Sequences of the N-terminal capping AR, the designed AR module and the C-terminal capping AR. The secondary structure elements are indicated above the sequences. The designed AR module consists of 26 defined framework residues, six randomized potential interaction residues (red x, any of the 20 natural amino acids except cysteine, glycine or proline) and one randomized framework residue (z, any of the amino acids asparagine, histidine or tyrosine). The designed AR module was derived via sequence and structure consensus analyses¹⁶. **(b)** Schematic representation of the library generation of designed AR proteins. Note that this assembly is represented on the protein level, whereas the real library assembly is on the DNA level. By assembling an N-terminal capping AR (green), varying numbers of the designed AR module (blue) and a C-terminal capping AR (cyan), combinatorial libraries of designed AR proteins of different repeat numbers were generated (side chains of the randomized potential interaction residues are shown in stick-mode in red). **(c)** Ribbon representation of the selected MBP binding AR protein off7 (colors as in **b**). This binder is derived from a library consisting of a N-terminal capping AR, three designed AR modules and a C-terminal capping AR. This figure was made with MolMol⁴⁹.

core (Fig. 1). We thereby increased the size of the potential interaction surface and potentiated its diversity. This strategy provided combinatorial libraries of designed AR proteins with distinct repeat numbers. The libraries were named N2C and N3C, indicating proteins had an N-terminal capping repeat, two and three, respectively, designed (and randomized) AR modules and a C-terminal capping repeat (Fig. 1). We used the N2C and N3C libraries for the selections, because AR proteins of this length are very abundant in nature⁶. Unselected members of these libraries were expressed in soluble form at about 200 mg/l in *E. coli* shake flask cultures. These proteins were monomeric, showed circular dichroism (CD) spectra indistinguishable from natural AR proteins¹⁶ and the AR fold was confirmed by a crystal structure of an unselected library member¹⁷. Similar to designed proteins in other consensus repeat protein studies^{18,19}, our designed AR proteins showed high thermodynamic stability during unfolding induced by heat¹⁶ and denaturants¹⁷. Hence, the consensus-designed AR proteins are stable scaffolds with large and modular potential interaction

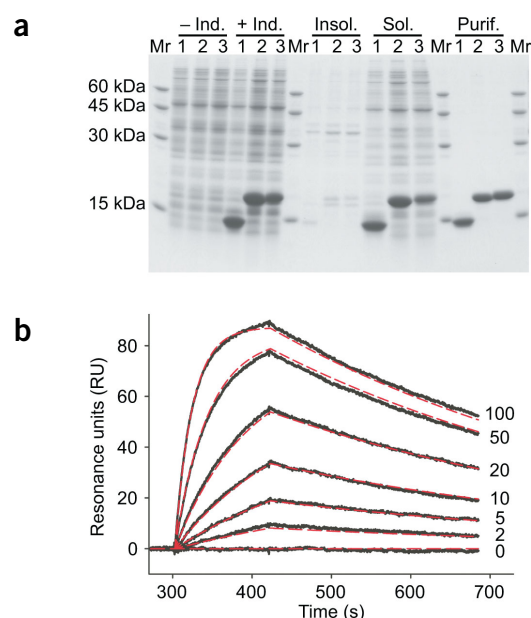


Figure 2 Expression, purification and SPR analysis of selected AR proteins. **(a)** Expression and purification of the selected MBP binders mbp3_16 (1), off7 (2) and mbp3_5 (3). At OD₆₀₀ = 0.6, the noninduced (– Ind.) cultures were induced with 0.5 mM IPTG and grown for 4 h at 37 °C (+ Ind.). After cell lysis, the AR proteins are in the soluble fraction (Sol.). Insol., insoluble fraction. The proteins were then purified in a single IMAC purification step (Purif.). **(b)** BIAcore analysis of off7. Different concentrations of off7 (0, 2, 5, 10, 20, 50 and 100 nM) were applied to a flow cell with immobilized MBP for 2 min, followed by washing with buffer flow. The global fit is indicated in the figure by red dashed lines (see Table 1 for the extracted kinetic data).

surfaces. The theoretical diversities of these libraries are $5.2 \cdot 10^{15}$ (N2C) and $3.8 \cdot 10^{23}$ (N3C)¹⁶. The DNA libraries used in the selections contained at least 10^{10} individual members each as estimated from the amount of ligated library DNA. The library diversities were further increased in subsequent PCR cycles.

Ribosome-display selection against MBP

We chose *E. coli* MBP as the first target protein for evaluating our libraries, because it can be obtained in large amounts in pure form and because its structure is known (Protein Data Bank (PDB) entry 1LLS)²⁰. We did the ribosome-display selections²¹ with biotinylated MBP bound to neutravidin in microtiter plates. An enrichment of binders was observed after the second selection round both for the N2C and the N3C libraries. We performed a total of four to five selection rounds before analyzing single, selected library members.

We screened individual selected AR proteins for MBP binding by an enzyme-linked immunosorbent assay (ELISA) using crude *E. coli* extracts. Of 60 N3C AR proteins screened, 18 gave a specific signal (signal/background ≥ 10), compared to 4 of 56 N2C molecules. Sequencing of the 18 MBP binding N3C AR proteins revealed that they could be divided into at least three sequence groups (see Supplementary Fig. 1 online). However, identical clones were never found and considerable diversity was left, indicating that an even more stringent selection pressure could be applied. For the N2C clones the sequence analysis was less conclusive because of the limited data set. Nevertheless, some repeats of N2C molecules showed striking sequence similarities to repeats of the selected N3C proteins. In both

Table 1 Kinetic binding data of selected clones determined by surface plasmon resonance

Target	Clone name (length)	k_{on} [$M^{-1}s^{-1}$]	k_{off} [s^{-1}]	K_D [M]
MBP	off7 (N3C)	$4.2 \cdot 10^5$	$1.9 \cdot 10^{-3}$	$4.4 \cdot 10^{-9}$
	mbp3_5 (N3C)	$2.0 \cdot 10^5$	$4.4 \cdot 10^{-3}$	$22 \cdot 10^{-9}$
	mbp3_16 (N2C)	$6.0 \cdot 10^5$	$1.0 \cdot 10^{-2}$	$17 \cdot 10^{-9}$
JNK2	JNK2_2_3 (N2C)	$9.7 \cdot 10^5$	$2.0 \cdot 10^{-3}$	$2.1 \cdot 10^{-9}$
p38	p38_2_3 (N2C)	$9.5 \cdot 10^5$	$3.5 \cdot 10^{-3}$	$3.7 \cdot 10^{-9}$

N2C and N3C molecules, aromatic residues appeared frequently and thus seem to be important in the selected sequences (see **Supplementary Fig. 1** online).

Framework mutations were randomly scattered and occurred at low frequencies but were present in all molecules (on average, 3.8 amino acid mutations per N3C and 2.4 mutations per N2C molecule; see **Supplementary Fig. 1** online). Each library member went through at least 450 PCR cycles (library generation and selection), which might explain this finding. The mutations are mostly of a conservative nature, and thus the fold of the AR domains is most probably not affected by the alteration of the framework (see description of the crystal structure below).

Selected AR proteins show high affinity and specificity

The selected AR proteins were expressed at high levels in soluble form in the cytoplasm of *E. coli* (up to 200 mg/l) and purified to homogeneity by a single immobilized metal ion affinity chromatography (IMAC) purification step (Fig. 2a). We screened 21 clones (4 N2C, 17 N3C) by surface plasmon resonance (SPR). Using the purified AR proteins at 1 μ M, we first compared the on- and off-rates of MBP binding. Three AR proteins with slow off-rates were analyzed at multiple concentrations and evaluated with a global kinetic fit (N2C: mbp3_16; N3C: off7 and mbp3_5) (Fig. 2b and Table 1). off7, a selected N3C molecule, had the highest affinity for MBP ($K_D = 4.4$ nM). The N2C molecule mbp3_16 had a dissociation constant of $K_D = 17$ nM. Hence, both N2C and N3C molecules can be selected to bind MBP with high affinity. Clones that went through five selection rounds had higher affinities for MBP than clones that were selected through four rounds. In SPR experiments, the N3C library member off7 was specific and did not cross-react with phage lambda protein D²², streptavidin or the aminoglycoside-3'-phosphotransferase APH(3')-IIIa, a bacterial kanamycin resistance protein²³.

To further investigate specificity, ELISA experiments were done with purified MBP-binding AR proteins (Fig. 3). mbp3_16, off7 and mbp3_5 are specific for MBP and do not interact with phage lambda protein D²², APH²³ or neutravidin (Fig. 3a). In a competition ELISA experiment, the binding of off7 to immobilized MBP could be inhibited by preincubation with free MBP (Fig. 3b). The affinity estimated from these experiments was consistent with the SPR measurements (50% inhibition at 10 nM). The unselected N3C AR library member E3_5^{16,17} did not interact with MBP, indicating that the designed AR domain scaffolds per se do not bind MBP (Fig. 3b).

At a concentration of 15 μ M, off7 is monomeric (as indicated by size exclusion chromatography) and shows a CD spectrum identical to that of E3_5 (ref. 16), which has an AR domain fold¹⁷ (data not shown).

Selection of specific high-affinity MAPK binders

To further evaluate the potential of our AR protein libraries, we chose the eukaryotic protein kinases JNK2 and p38 as our next target

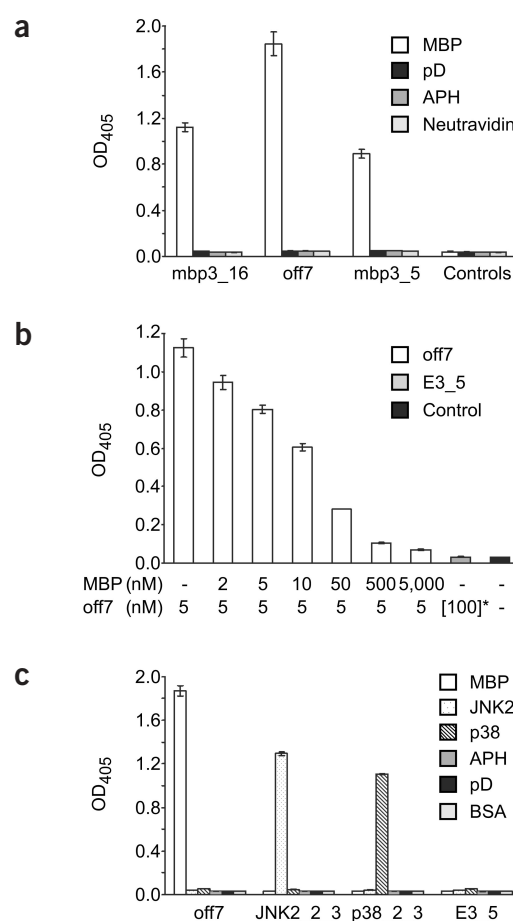


Figure 3 ELISAs with selected AR proteins. (a) Specificity of MBP binders. The interaction of the proteins mbp3_16, off7 and mbp3_5 (each 50 nM; control with no AR protein) with immobilized MBP, pD, APH and neutravidin is shown. (b) Competition ELISA illustrating the interaction between the selected AR protein off7 and MBP. off7 (5 nM) was incubated with varying concentrations of free MBP before binding on immobilized MBP. The binding to MBP of off7 can be specifically inhibited by increasing concentrations of free MBP in solution. An unselected AR protein of the N3C library (*) showed no interaction with MBP (100 nM of E3_5)^{16,17}, giving a signal identical to that of the control (no AR protein on immobilized MBP). (c) Specificity comparison of an MBP, a JNK2 and a p38 binder. The interaction of 100 nM each of the proteins off7 (binds MBP), JNK2_2_3 (binds JNK2), p38_2_3 (binds p38) and E3_5 (unselected N3C library member) with MBP, JNK2, p38, APH, pD and BSA is shown. Note that in all representations the background binding of the detection antibodies has not been subtracted.

proteins (see ref. 24 and references therein). We did a total of four ribosome-display selection rounds with the N2C library before comparing single, selected library members. Screening 15 clones each, we obtained ten ELISA-positive JNK2 binders and ten ELISA-positive p38 binders. The sequence of one representative member for each of these target-specific groups is given in **Supplementary Figure 1** online. These MAPK binders share many features of the selected MBP binders. They have affinities in the low nM range (Table 1); they can be expressed at high levels in soluble form in the cytoplasm of *E. coli* and purified to homogeneity by a single IMAC purification step (data not shown); their randomized positions are enriched in aromatic amino acids (see **Supplementary Fig. 1** online). To investigate their target specificity, we did ELISA experiments with purified, selected AR

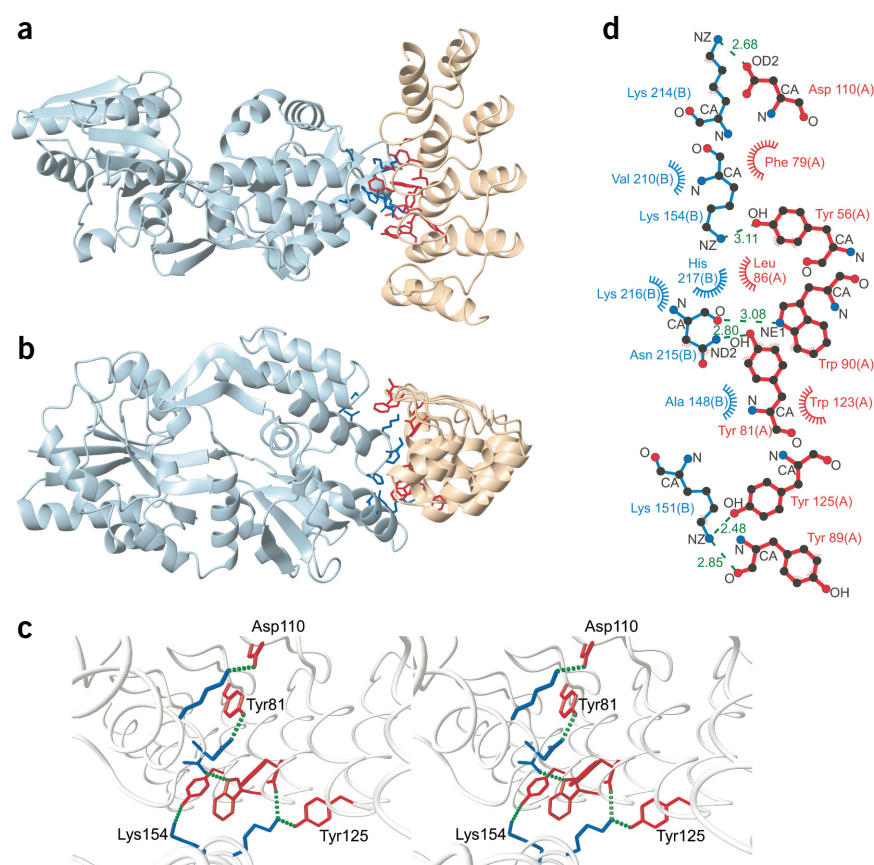


Figure 4 Crystal structure of the designed AR protein off7 in complex with MBP. **(a,b)** Two perpendicular views of the complex are shown. MBP is on the left (blue), off7 on the right (ochre). The interaction residues are highlighted in stick-mode in red (off7) and blue (MBP), respectively. **(c)** A close stereo view on the H-bond pattern in similar view as in **b**. Note that in this representation only residues involved in H-bonds (green dashed lines) are shown. For orientation, some residues involved in H-bonding are labeled. **Figure 4a–c** were made with MolMol⁴⁹. **(d)** Ligplot⁴⁸ representation of the interaction between MBP (chain B, blue) and off7 (chain A, red). H-bonds (in green) including the H-bond distances as well as residues and atoms involved in hydrophobic contacts (indicated by red or blue rays) are shown.

proteins (Fig. 3c). All binders were highly specific for the target proteins that were used to select them. Most importantly, the AR proteins allow perfect discrimination between the homologous MAPKs (51% identity and 59% similarity between JNK2 and p38 on the amino acid level in our format).

Structure determination of off7 in complex with MBP

To validate our AR randomization scheme and to analyze the selected interaction at the atomic level, we determined the crystal structure of one binder (off7) in complex with MBP (see Methods). The phasing problem was solved by molecular replacement using E3_5, a designed N3C AR protein¹⁷ as a search model without using the phases of the larger MBP. Hence, when used in cocrystallization studies, AR proteins may serve as valuable tools to obtain first phases and finally the structure of its binding partner. The results of the data collection and refinement are shown in **Supplementary Table 1** online. As can be seen in the crystal structure (Figs. 4 and 5), the AR protein binds the open form of MBP²⁰ creating an elongated complex. For the AR protein, clear electron density starts at Ser12, but no or only very weak electron density was observed for the N-terminal His₆ tag. The electron density

is clear until the second-to-last amino acid (Leu168). In the complex, the AR protein has the typical AR domain fold and is highly similar to the known structure of the designed AR protein E3_5 (root-mean-square deviation of the C_α atoms (r.m.s.d._{Cα}) < 1 Å)¹⁷. The main differences between the two structures are found in the β-turn region of repeat module 1 (second repeat), where the C_α chain shows a maximal r.m.s.d._{Cα} of 2.1 Å. The H-bond network, which most probably stabilizes the β-turn region¹⁷, is conserved and undisturbed. off7 has three framework mutations (K16R, N74D and H125Y) not present in the library design, but they are conservative and in accordance with the AR domain fold, and only Y125 is involved in MBP binding (see below).

MBP was found in the open conformation with no ligand bound²⁰. A superposition on PDB entry 1LLS²⁰ shows very few differences (r.m.s.d._{Cα} < 0.9 Å). The N-terminal His₆ tag and the C terminus of MBP are not defined in the electron density. Clear density extends from Gly19 to Gly387.

Analysis of the interaction of off7 with MBP

The interaction of the AR protein off7 with MBP was analyzed as described in the Methods section. The only direct interaction between off7 and MBP in the crystal lattice is the selected heterodimer interface (Fig. 4). Crystal packing contacts are mainly between adjacent off7 molecules and between adjacent MBP molecules. The heterodimer interface is formed by the concave randomized surface of the AR protein off7 (611 Å² buried surface) and a slightly larger convex surface on the MBP (656 Å² buried surface), resulting in a total buried surface area of 1,267 Å² (Fig. 5).

The off7/MBP complex is further characterized by six H-bonds and a planarity index of 2.1²⁵. The details of the interaction are listed in **Table 2** and in **Supplementary Table 2** online.

The interaction of off7 involves residues from all three randomized repeat modules, although the randomized repeat modules 2 and 3 (constituting repeats 3 and 4 in the protein because of the capping repeats) contribute more to the binding than the randomized repeat module 1 does. In total, 9 out of 18 randomized potential interaction residues are involved in the binding to MBP. The interface is characterized by a large number of aromatic residues, which account for 73% of the buried surface area on off7 (Fig. 4 and **Supplementary Table 2** online). Among these residues, the tyrosines cover 28% of the buried surface area and are involved in four H-bonds (Tyr56, Tyr81, Tyr89 and Tyr125). In the interface (Fig. 4), tyrosines have a dual role being both H-bond formers and hydrophobic contact mediators. Besides the four tyrosines, Trp90 and Asp110 also form H-bonds. Three framework residues (Leu86, Asp110 and Tyr125) form part of the interaction surface. Two (Asp110 and Tyr125) form H-bonds to MBP (Fig. 4) and the third (Leu86) is engaged in hydrophobic interactions. Interestingly, Tyr125 is a framework mutation (H125Y). With the

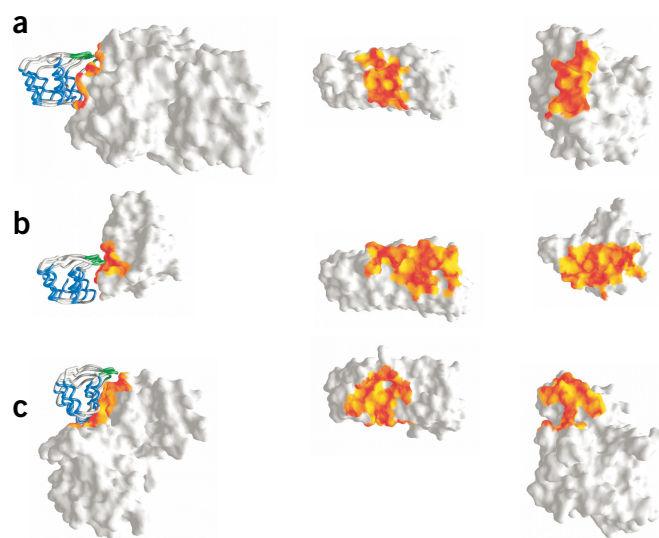


Figure 5 Open sandwich illustrations of the interaction surfaces of AR proteins and their targets. (a–c) GRASP⁵⁰ shape complementarity representations of the interactions between off7 and MBP (shape complementarity, 0.739) (a), GABPβ1 and GABPα (PDB entry 1AWC; shape complementarity, 0.665) (b), and p18^{INK4c} and CDK6 (PDB entry 1G3N; shape complementarity, 0.688) (c), respectively. The complex is shown on the left with the AR proteins in a backbone worm representation (α-helices in blue, β-turns in green) and the target protein in a surface representation. The open sandwich surface representations are shown in the middle (AR proteins) and on the right (targets). The contact areas are stained according to the shape complementarities from orange (low) to red (high).

exception of the high percentage of aromatic amino acids, the residue composition of the off7 interface is comparable to other AR protein complexes. However, the small and functionally biased set of only seven structurally analyzed natural AR protein complexes limits the comparison to their average amino acid composition in the binding interface.

The interaction surface of MBP is located on helices H7, H9 and H17 (Fig. 4 and **Supplementary Table 2** online). Four lysines (Lys151, Lys154, Lys214 and Lys216) are involved in four H-bonds and together form approximately 60% of the buried surface area on the MBP surface (Fig. 4). Every lysine of MBP is in contact with a tyrosine of off7 resulting in three H-bonds and five hydrophobic contacts (see **Supplementary Table 2** online). In general, the surface of the MBP is rather negatively charged except for the spot where off7 binds, where there is a positively charged surface patch formed by the four lysines.

A comparison to natural protein-protein interactions

We compared the off7/MBP complex to all available complex crystal structures of AR proteins, to natural heterodimer complexes, to antibody-antigen complexes and to an affibody complex (see Methods section and **Table 2**). The off7/MBP interaction is comparable to natural protein-protein interactions²⁵. The 611 Å² buried surface area, one H-bond per 100 Å² buried surface area and the planarity value of 2.1 of off7, are all within the normal parameters of heterodimer complexes (**Table 2**)²⁵. The buried surface area is at the lower limit of antibody-antigen complexes²⁶, whereas the affinity is comparable to high affinity monovalent antibody-antigen binding. Otherwise, all values seem to fit the standard parameters for antibody-antigen complexes quite well, but are also similar to those of heterodimer complexes. Natural AR protein interactions can be very diverse in size,

composition and in the structural elements involved (Fig. 5 and **Table 2**). Our randomization scheme of the β-turn and the first α-helix of the designed AR modules was based on crystal structures of natural AR complexes, where it was clear that the AR scaffold is directly used for binding¹⁶. Because 50% of the randomized target interaction residues that were selected do indeed interact with MBP, the interaction mode is very similar to that of natural AR proteins, such as the GABPβ1 (Fig. 5), which partly inspired our library design¹⁶. In comparison to the natural AR proteins, off7 has a slightly smaller buried surface area but a higher H-bond density (**Table 2**).

Recently the structure of an affibody, another designed binding molecule, based on the staphylococcal three-helix bundle protein A, in complex with its target protein was published²⁷. The complex shows a slightly larger buried surface area than the off7/MBP complex but with a comparable number of H-bonds (**Table 2**). The affibody has a thousand-fold lower affinity for its target²⁸ than the AR protein off7 does for MBP ($K_D = 6 \mu\text{M}$ vs. $K_D = 4.4 \text{ nM}$, respectively). This is probably because this affibody is in a molten globule state and assumes a defined structure only upon binding²⁹, leading to a loss of entropy that reduces the overall free energy of binding and thus the observed affinity.

DISCUSSION

We designed AR protein libraries of varying repeat numbers using a consensus design strategy^{15–17}. Here we show the successful selection of binding molecules from these libraries. The properties of the designed AR proteins perfectly match the criteria for alternative scaffolds. They are expressed at a high level in soluble form, are monomeric and do not contain any cysteines¹⁶. Unlike previously presented scaffolds (for reviews see refs. 12,13), which typically use either randomized loops or a randomized surface on a given protein scaffold for binding, AR proteins use both β-turns and a randomized surface. Most importantly, they are not restricted in dimension because of their modular architecture (Fig. 1). Thus, the interaction surface can be adapted by adding more repeat modules. The favorable properties of the molecules in the starting libraries seem to positively influence both the speed of selection and its outcome. In only four to five ribosome-display selection rounds, we were able to enrich the pool in specific, high-affinity protein binders from the N2C and N3C libraries, which appears to be faster than ribosome-display selections from antibody single-chain Fv libraries³⁰.

The selected AR proteins retain the advantageous properties of the designed AR proteins, being expressed in high amounts in soluble form and free of cysteines (Fig. 2). The selected clones specifically recognize the target protein against which they were selected and do not cross-react with other proteins as shown by ELISA (Fig. 3). The affinities of the selected clones are in the low nanomolar range (**Table 1** and Fig. 2), the association rates are in the typical range for protein-protein interactions (that is, 10^5 – $10^6 \text{ M}^{-1}\text{s}^{-1}$)³¹ and the dissociation rates are in the range of 10^{-2} to $2 \cdot 10^{-3} \text{ s}^{-1}$ (**Table 1**). We anticipate that these off-rates can probably be improved further by a diversification step followed by an off-rate ribosome-display selection round³². Such a diversification step could involve not only classical error-prone PCR, but also, more interestingly, strategies purely amenable to repeat proteins such as repeat shuffling or repeat elongation.

The selected AR protein sequences contain a high number of aromatic residues (see **Supplementary Fig. 1** online) and in the case of off7, seven aromatic amino acids are involved in MBP binding (Fig. 4), including four prominent tyrosines (**Supplementary Table 2** online). Interestingly, high tyrosine content has also been noted in antibody binding sites²⁶. The dual interaction role of tyrosine as H-bond former and hydrophobic contact mediator (see Results)³³ is probably the

Table 2 Comparison between AR protein complexes and other protein-protein interactions

PDB entry	Resolution	ΔASA^a in $\text{\AA}^2$	No. H-bonds	No. H-bonds/100 $\text{\AA}^2$ ΔASA^a	No. of salt bridges	Planarity	No. of bridging H_2O
1awc	2.2	853.62	5	0.58	0	2.30	3
1bi7	3.4	1205.5	7	0.29	4	2.50	0
1blx	1.9	845.4	11	1.30	1	2.50	12
1g3n	2.9	843.3	12	1.42	1	2.20	0
1ikn	2.3	763.96	2	0.26	n.d.	2.73	0
1oy3	2.1	1593.9	7	0.44	n.d.	4.01	10
1ycs ^b	2.2	713.7	7	1.00	1	2.60	1
1svx (off7/MBP)	2.3	611.2	6	1.00	0	2.10	0
1lp1 (Affibody)	2.3	848.9	6	0.71	0	2.00	2
Antibody-antigen complexes ^c	—	777 $\pm$ 135	—	1.1 $\pm$ 0.5	—	2.2 $\pm$ 0.4	—
Heterodimeric protein-protein complexes ^c	—	983 $\pm$ 582	—	1.1 $\pm$ 0.5	—	2.8 $\pm$ 0.9	—

^aSurface area per molecule occluded upon complex formation. ^b1ycs uses a different binding surface than the other AR proteins in this table. ^cAccording to ref. 25.

reason for this accumulation. In addition, aromatic residues are generally enriched in protein-protein interaction interfaces²⁵.

Apart from the high content of aromatic residues, the crystal structure of the off7/MBP complex reveals an interaction interface that is comparable to that found in natural heterodimer and antibody-antigen complexes (Table 2)²⁵. It also shows that the AR protein binds its target with the randomized amino acids (Fig. 4 and Supplementary Table 2 online), hence validating our randomization scheme. In the crystal structure of the complex, off7 shows the typical AR domain fold with a high similarity to the unselected N3C library member E3_5 (ref. 17). In solution, the uncomplexed, monomeric off7 shows a CD spectrum virtually identical to that of E3_5 (ref. 16). Thus, the backbone of off7 does not seem to be rearranged in a substantial way upon binding to MBP but rather seems to interact in a key-to-lock mechanism. Such a rigid-body interaction may be advantageous both for affinity (low entropic costs upon binding) and specificity (conformational restriction). p18, a natural AR protein interacting with CDK4/6 (Fig. 5), also has a low r.m.s.d._{C α} of < 0.9 $\text{\AA}$ between the complexed (PDB entry 1G3N) and the uncomplexed (PDB entry 1IHB) AR protein. However, the interaction with CDK4/6 leads to structural alterations in the target³⁴. In contrast to those of the rigid AR domain scaffolds, the binding site of antibodies seem to be able to adopt different conformations. The loops of the complementarity-determining region of typical protein binding antibodies can undergo substantial changes upon binding³⁵ and the loop flexibility might even be used to accomplish multispecificity³⁶.

Here, we have successfully validated our AR proteins as designed binding molecules using MBP and two MAPKs as model targets, and thus have introduced a binding molecule with very favorable properties. The combination of the high expression level of designed AR proteins (200 mg/l soluble protein in shake flasks), their high thermodynamic stability (9.5 to 21 kcal/mol¹⁷), the absence of cysteines, the fast enrichment of binders (four selection rounds), their low nanomolar affinities along with high specificities and their modular architecture compares favorably with reports on antibodies and other alternative scaffolds (such as protein A²⁸, lipocalins¹⁴, fibronectin³⁷ or green fluorescent protein³⁸). Our results open the door for a number of applications. Apart from being useful in capturing molecules on protein chips or in affinity purification—typical applications for designed binding molecules—designed AR proteins are especially suited for intracellular applications. Their high stability, the absence of disulfide bonds and the selectable high affinities are optimal prerequisites for intracellular inhibitors³⁹, where antibodies are less than ideal.

Generally, AR proteins are also conceivable in therapeutic applications, which are currently a domain of recombinant antibodies. As we have shown here, it is possible to both cocrystallize a target protein with a designed binding AR protein and determine its crystal structure with the help of the AR protein. Hence, designed binding AR proteins could be used in cocrystallization and structure determination of proteins difficult to crystallize, similar to what has been shown with antibodies⁴⁰.

METHODS

Molecular biology. Unless stated otherwise, all experiments were done according to protocols found in reference 41. Enzymes and buffers were from New England Biolabs (NEB) or Fermentas. All PCR reactions were done using the proofreading Vent-polymerase (NEB).

Vectors used in antigen production. The different vectors that were prepared for the present study are described in detail in the **Supplementary Methods** online. pQEMBP (GenBank accession no. AY327141) was used for the expression of His-tagged, nonbiotinylated MBP. pAT224 (AY327139) was used for the expression of His-tagged, biotinylated MBP. pAT222 (AY327137) was used for the production of His-tagged, biotinylated pD. pAT222_JNK2 and pAT222_p38 were used for the production of His-tagged, biotinylated JNK2 and p38, respectively. All pAT222 and pAT224 constructs carry an Avi tag for biotinylation at the N terminus and a His₆ tag at the C terminus. pBirAcm (Avidity) was used for *in vivo* biotinylation.

Antigen production and purification. The biotinylated proteins pD, MBP, JNK2 and p38 (plasmids pAT222, pAT224, pAT222_JNK2 and pAT222_p38) were produced using *in vivo* biotinylation with plasmid pBirAcm in *E. coli* XL-1 Blue (Stratagene) according to the protocols of Avidity and QIAGEN. Efficient biotinylation was confirmed by ELISA and blotting with a streptavidin-alkaline phosphatase conjugate (Roche) and mass spectrometry. Nonbiotinylated MBP for the ELISA analysis and crystallization was produced in the same way as the AR proteins¹⁶ using pQEMBP in *E. coli* XL-1 Blue. The protein purification was carried out as described¹⁶.

Ribosome-display vector (pRDV; AY327136). The cloning of pRDV is described in detail in the **Supplementary Methods** online. pRDV contains all flanking DNA regions necessary for ribosome display: the T7-promoter, the ribosomal binding site and an in-frame *tolA* gene spacer. Hence, by simple ligation of the DNA encoding the combinatorial library into pRDV and by a PCR using this ligation mix as template, all features necessary for ribosome display are added to the library. The use of pRDV has the advantage that it always provides error-free library flanking regions and that it saves a number of working steps compared to the standard PCR approach for library generation²¹.

Generation of combinatorial libraries. The AR protein library generation has been described¹⁶. We changed that protocol in this study in that all ARs, that is,

the N-terminal capping AR, the designed repeat module and the C-terminal capping AR were used as PCR products for the assembly of the libraries. In the present study, both the N- and the C-terminal capping repeat were amplified by PCR from cloned and verified sequences to reduce sequence errors. In this way, AR protein libraries consisting of an N-terminal capping AR, two or three designed AR modules and a C-terminal capping AR (N2C and N3C libraries) were assembled. To convert the libraries to the ribosome-display format, they were amplified by PCR using oligonucleotides EWT4 (5'-TTCCTCATGAGAGATCGCATCACCATCACCATCACGGATCCGACCTGGG-3') and WTC4 (5'-TTTGGGAAGCTTTTGACGATTTCAGC-3') and ligated into pRDV using the restriction enzymes *Bsp*HI (or *Nco*I for pRDV) and *Hind*III. The ligation product was purified using QIAquick (QIAGEN) columns. The purified ligation served as template for a PCR using oligonucleotides RDVf1 (5'-CCTTTTGCTCACATGACCCG-3') and tolAk (5'-CCGCACACCAAGGTGTGCGGTTTCAGTTGCCGCTTTCTTTCT-3'). Thereby, combinatorial N2C and N3C DNA libraries were generated in the ribosome-display format.

Ribosome display. The PCR-amplified libraries were transcribed and selections were done as described²¹. For the selection, the biotinylated antigen was immobilized as follows: neutravidin (66 nM, 100 µl/well; Pierce) in TBS150 (50 mM Tris HCl, pH 7.4, 150 mM NaCl) was immobilized on a Maxisorp plate (Nunc) by overnight incubation at 4 °C. The wells were then blocked with 300 µl 0.5% BSA (Fluka) in TBS150 for 1 h at 23 °C. Biotinylated antigen (100 µl, 1 µM) in TBS150 with 0.5% BSA was allowed to bind for 1 h at 4 °C. Before the ribosome-display round, the wells were extensively washed with washing buffer WBT (50 mM Tris acetic acid, pH 7.5, 150 mM NaCl, 50 mM Mg(CH₃COO)₂, 0.05% Tween 20). A ribosome-display round consisted of two 30-min pre-panning steps on neutravidin and a 1 h binding step on the target protein. After washing, RNA purification and reverse transcription (with oligonucleotide tolAk), a first PCR was done using oligonucleotides T7B (5'-ATACGAAATTAATACGACTCACTATAGGGAGACCACAACGG-3') and tolAk. This RT-PCR product was purified on an agarose gel and reamplified in a second PCR using the same oligonucleotides. The second PCR product served as template for the next round of ribosome display. The number of RT-PCR cycles was reduced from 40 to 30 to 25 in the first three rounds to monitor the enrichment of binders. Binders were analyzed after four or five rounds.

Analysis of selected binders. From the selected DNA pools, the AR open reading frame was amplified by PCR and cloned into pQE30 (QIAGEN) via *Bam*HI/*Hind*III (oligonucleotides: EWT3: 5'-TTCCGCGGATCCGACCTGGG-3' and WTC4). The DNA sequences were determined using standard techniques. The sequences of the MBP binding proteins off7 (AY326424), MBP3_5 (AY326425) and MBP3_16 (AY326426) have been deposited in GenBank. The amino acid sequences of all sequenced clones are listed in **Supplementary Figure 1** online. For the ELISA screening, the crude extract of 0.6 ml protein expression cultures was used (expression according to QIAGEN). The cell pellets were lysed with 50 µl B-Per (Pierce) and the lysates were mixed with 250 µl TBS500 (50 mM Tris HCl, pH 8.0, 500 mM NaCl) each. For quantitative ELISA, BIAcore, CD, analytical gel-filtration and crystallization, single, selected library members were produced on a liter scale and purified as described¹⁶. CD spectroscopy and analytical gel-filtration were done as described¹⁶ using 15 µM protein in TBS150 (pH 7.4).

ELISA. Biotinylated antigens were immobilized on neutravidin-coated plates as described above. For the screening of the pools, 100 µl of the above crude extracts were applied to wells with or without immobilized antigen for 1 h at 4 °C. After extensive washing with TBS150, binding was detected with an anti-RGS-His antibody (QIAGEN; detects only the RGS-His₆-tag of the AR protein, not the His₆-tag of the antigen), an anti-mouse-IgG-alkaline phosphatase conjugate (Pierce) and p-nitrophenylphosphate (Fluka). Quantitative ELISAs were done in the same manner, except purified protein was used (see **Fig. 3**). For competition ELISA, the purified AR protein off7 was incubated with varying amounts of free MBP before (4 °C, 100 min) and during the binding reaction (see **Fig. 3**).

Surface plasmon resonance (SPR). SPR was measured using a BIAcore 3000 instrument (BIAcore). The running buffer was 20 mM HEPES, pH 7.4, 150 mM NaCl and 0.005% Tween 20. A streptavidin SA chip (BIAcore) was

used with 480 RU biotinylated MBP immobilized (440 RU JNK2 and 450 RU p38, respectively). The interactions were measured at a flow of 60 µl/min with 5 min buffer flow, 2 min injection of MBP-binding AR protein in varying concentrations (10 pM to 200 nM) and an off-rate measurement of 40 min with buffer flow. The signal of an uncoated reference cell was subtracted from the measurements. Inhibition BIAcore measurements gave results similar to that of the kinetic analyses (data not shown). The p38 and JNK2 binders were measured similarly, but with an injection time of 3 min. The kinetic data of the interaction were evaluated with a global fit using BIAevaluation 3.0 (BIAcore), Scrubber (BioLogic software) and Clamp⁴².

Complex purification and crystallization. MBP and the selected MBP-binding AR proteins off7, mbp3_5 and mbp3_16 were produced as described above. The cell pellets of 1-liter bacterial culture of each MBP and off7 (or mbp3_5 or mbp3_16) were pooled and then lysed using an Emulsiflex C5 (Avestin) followed by additional sonication. The proteins were purified using an IMAC column as described¹⁶, followed by a preparative Superdex-75 (Amersham Pharmacia) size exclusion chromatography step in 10 mM Tris HCl, pH 7.6 and 100 mM NaCl. For every protein mixture, the peak fraction with the smallest molecular weight containing both MBP and the AR protein in equimolar amounts, as determined by SDS-PAGE, was collected and used for crystallization. Light scattering of these fractions was measured as described¹⁷. It showed a monodisperse particle distribution for all three complexes. For the off7/MBP complex (61.4 kDa calculated mass for the 1:1 complex) an average radius of 3.6 nm, equivalent to a hydrated particle of 95 kDa, was estimated, which corresponds to a nonhydrated particle of 65 kDa. The off7/MBP complex crystallized readily and was further analyzed. The protein complex was concentrated to 26 mg/ml for crystallization. Initial crystallization screening was done in 96-well, sitting drop, square well crystallization plates (Greiner Bio-One). The reservoirs were filled with 100 µl reservoir solution using an 8-channel pipette from a 2 ml 96-well, deep-well block into the crystallization plate. Using an 8-channel pipette, 2 µl of reservoir solution were pipetted in the crystallization well and mixed with 2 µl of protein solution. The initial crystals were refined using standard techniques. The crystals used for data collection grew in about 2–3 weeks in 30% PEG 6000, 0.1 M Tris HCl pH 8–9, 100 mM NaCl in a hanging drop experiment with 500 µl reservoir, 2 µl protein solution mixed with 2 µl water and 2 µl reservoir solution. For data collection the crystals were soaked in the mother liquid with 10% ethylene glycol for about 30 s to 1 min and flash frozen in a cryostream at 100 K.

Data collection, reduction, structure solution and refinement. Data were collected at the European Synchrotron Radiation Facility beamline ID14-1. The data were processed using MOSFLM, SCALA and TRUNCATE⁴³. The crystal belonged to space group P2₁, with a Matthews coefficient of $V_M = 2.1 \text{ Å}^3/\text{Da}$, corresponding to an estimated water content of 39%.

The crystal structure was determined by molecular replacement using the program AmoRe⁴⁴, with the structure of the unselected N3C library member E3_5 (PDB entry 1MJ0¹⁷) as a search model. A conventional AmoRe protocol (rotation, translation, rigid body refinement) was applied yielding a solution. This information was used to obtain a first electron density. At this point only the AR protein was clearly visible in the electron density. A solvent flipping protocol was then applied to modify the map⁴⁵. MBP in its open form (PDB entry 1LLS)²⁰ was positioned in the resulting electron density using program O⁴⁶. Because about a third of MBP was visible, manual building would have been feasible as well. The rest of the model building was carried out using the program O⁴⁶, the structure refinement was done in CNS⁴⁵ resulting in a final model with an R-factor of 19.5 % and an R_{free}-factor of 24.9 % (**Supplementary Table 1** online).

Analysis of the complexes. The structural analysis of the complexes was done as suggested²⁵. H-bonds and hydrophobic interactions were calculated with HBPLUS⁴⁷, LIGPLOT⁴⁸ and DIMPLOT⁴⁸ using the default settings. Other parameters were calculated using the protein-protein interaction server <http://www.biochem.ucl.ac.uk/bsm/PP/server/index.html>²⁵ or CNS⁴⁵.

The atomic coordinates of the described complex were deposited in the PDB (PDB-ID: 1SVX).

Note: Supplementary information is available on the Nature Biotechnology website.

ACKNOWLEDGMENTS

We thank the members of the Plückthun and Grütter laboratories for valuable discussions and David L. Zechel for the critical reading of the manuscript. H.K.B. was supported by a pre-doctoral fellowship of the Roche Research Foundation. M.T.S. was in receipt of a pre-doctoral scholarship from the Fonds der Chemischen Industrie and the Bundesministerium für Bildung und Forschung. This work was supported by the Swiss National Center of Competence in Research in structural biology and the Swiss Krebsliga.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests (see the *Nature Biotechnology* website for details).

Received 24 December 2003; accepted 14 January 2004

Published online at <http://www.nature.com/naturebiotechnology/>

1. Groves, M.R. & Barford, D. Topological characteristics of helical repeat proteins. *Curr. Opin. Struct. Biol.* **9**, 383–389 (1999).
2. Andrade, M.A., Perez-Iratxeta, C. & Ponting, C.P. Protein repeats: structures, functions, and evolution. *J. Struct. Biol.* **134**, 117–131 (2001).
3. Kobe, B. & Kajava, A.V. When protein folding is simplified to protein coiling: the continuum of solenoid protein structures. *Trends Biochem. Sci.* **25**, 509–515 (2000).
4. Marcotte, E.M., Pellegrini, M., Yeates, T.O. & Eisenberg, D. A census of protein repeats. *J. Mol. Biol.* **293**, 151–160 (1999).
5. Ellis, J., Dodds, P. & Pryor, T. Structure, function and evolution of plant disease resistance genes. *Curr. Opin. Plant Biol.* **3**, 278–284 (2000).
6. Bork, P. Hundreds of ankyrin-like repeats in functionally diverse proteins: mobile modules that cross phyla horizontally? *Proteins: Struct. Funct. Genet.* **17**, 363–374 (1993).
7. Sedgwick, S.G. & Smerdon, S.J. The ankyrin repeat: a diversity of interactions on a common structural framework. *Trends Biochem. Sci.* **24**, 311–316 (1999).
8. Suzuki, F. *et al.* Functional interactions of transcription factor human GA-binding protein subunits. *J. Biol. Chem.* **273**, 29302–29308 (1998).
9. Malek, S., Huxford, T. & Ghosh, G. Ix α functions through direct contacts with the nuclear localization signals and the DNA binding sequences of NF- κ B. *J. Biol. Chem.* **273**, 25427–25435 (1998).
10. Winter, G. & Milstein, C. Man-made antibodies. *Nature* **349**, 293–299 (1991).
11. Plückthun, A. *et al.* in *Antibody Engineering* (eds. McCafferty, J., Hoogenboom, H.J. & Chiswell, D.J.) 203–252 (IRL Press, Oxford, 1996).
12. Nygren, P.-Å. & Uhlen, M. Scaffolds for engineering novel binding sites in proteins. *Curr. Opin. Struct. Biol.* **7**, 463–469 (1997).
13. Skerra, A. Engineered protein scaffolds for molecular recognition. *J. Mol. Recognit.* **13**, 167–187 (2000).
14. Skerra, A. Imitating the humoral immune response. *Curr. Opin. Chem. Biol.* **7**, 683–693 (2003).
15. Forrer, P., Stumpp, M.T., Binz, H.K. & Plückthun, A. A novel strategy to design binding molecules harnessing the modular nature of repeat proteins. *FEBS Lett.* **539**, 2–6 (2003).
16. Binz, H.K., Stumpp, M.T., Forrer, P., Amstutz, P. & Plückthun, A. Designing repeat proteins: well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins. *J. Mol. Biol.* **332**, 489–503 (2003).
17. Kohl, A. *et al.* Designed to be stable: crystal structure of a consensus ankyrin repeat protein. *Proc. Natl. Acad. Sci. USA* **100**, 1700–1705 (2003).
18. Mosavi, L.K., Minor, D.L. Jr. & Peng, Z.-y. Consensus-derived structural determinants of the ankyrin repeat motif. *Proc. Natl. Acad. Sci. USA* **99**, 16029–16034 (2002).
19. Main, E.R., Xiong, Y., Cocco, M.J., D'Andrea, L. & Regan, L. Design of stable alpha-helical arrays from an idealized TPR motif. *Structure (Camb)* **11**, 497–508 (2003).
20. Rubin, S.M., Lee, S.Y., Ruiz, E.J., Pines, A. & Wemmer, D.E. Detection and characterization of xenon-binding sites in proteins by ^{129}Xe NMR spectroscopy. *J. Mol. Biol.* **322**, 425–440 (2002).
21. Hanes, J. & Plückthun, A. *In vitro* selection and evolution of functional proteins by using ribosome display. *Proc. Natl. Acad. Sci. USA* **94**, 4937–4942 (1997).
22. Yang, F. *et al.* Novel fold and capsid-binding properties of the lambda-phage display platform protein gpD. *Nat. Struct. Biol.* **7**, 230–237 (2000).
23. Hon, W.C. *et al.* Structure of an enzyme required for aminoglycoside antibiotic resistance reveals homology to eukaryotic protein kinases. *Cell* **89**, 887–895 (1997).
24. Forrer, P., Tamaskovic, R. & Jaussi, R. Enzyme-linked immunosorbent assay for the measurement of JNK, ERK and p38 kinase activities. *Biol. Chem.* **379**, 1101–1111 (1998).
25. Jones, S. & Thornton, J.M. Principles of protein-protein interactions. *Proc. Natl. Acad. Sci. USA* **93**, 13–20 (1996).
26. Lo Conte, L., Chothia, C. & Janin, J. The atomic structure of protein-protein recognition sites. *J. Mol. Biol.* **285**, 2177–2198 (1999).
27. Högbom, M., Eklund, M., Nygren, P.-Å. & Nordlund, P. Structural basis for recognition by an *in vitro* evolved affibody. *Proc. Natl. Acad. Sci. USA* **100**, 3191–3196 (2003).
28. Eklund, M., Axelsson, L., Uhlen, M. & Nygren, P.-Å. Anti-idiotypic protein domains selected from protein A-based affibody libraries. *Proteins: Struct. Funct. Genet.* **48**, 454–462 (2002).
29. Wahlberg, E. *et al.* An affibody in complex with a target protein: structure and coupled folding. *Proc. Natl. Acad. Sci. USA* **100**, 3185–3190 (2003).
30. Hanes, J., Schaffitzel, C., Knappik, A. & Plückthun, A. Picomolar affinity antibodies from a fully synthetic naive library selected and evolved by ribosome display. *Nat. Biotechnol.* **18**, 1287–1292 (2000).
31. Wodak, S.J. & Janin, J. Structural basis of macromolecular recognition. *Adv. Protein Chem.* **61**, 9–73 (2002).
32. Jermutus, L., Honegger, A., Schwesinger, F., Hanes, J. & Plückthun, A. Tailoring *in vitro* evolution for protein affinity or stability. *Proc. Natl. Acad. Sci. USA* **98**, 75–80 (2001).
33. Mian, I.S., Bradwell, A.R. & Olson, A.J. Structure, function and properties of antibody binding sites. *J. Mol. Biol.* **217**, 133–151 (1991).
34. Jeffrey, P.D., Tong, L. & Pavletich, N.P. Structural basis of inhibition of CDK-cyclin complexes by INK4 inhibitors. *Genes Dev.* **14**, 3115–3125 (2000).
35. Sundberg, E.J. & Mariuzza, R.A. Molecular recognition in antibody-antigen complexes. *Adv. Protein Chem.* **61**, 119–160 (2002).
36. James, L.C., Roversi, P. & Tawfik, D.S. Antibody multispecificity mediated by conformational diversity. *Science* **299**, 1362–1367 (2003).
37. Xu, L. *et al.* Directed evolution of high-affinity antibody mimics using mRNA display. *Chem. Biol.* **9**, 933–942 (2002).
38. Zeytun, A., Jeromin, A., Scalettar, B.A., Waldo, G.S. & Bradbury, A.R. Fluorobodies combine GFP fluorescence with the binding characteristics of antibodies. *Nat. Biotechnol.* **21**, 1473–1479 (2003).
39. Cattaneo, A. & Biocca, S. The selection of intracellular antibodies. *Trends Biotechnol.* **17**, 115–121 (1999).
40. Ostermeier, C., Iwata, S., Ludwig, B. & Michel, H. Fv fragment-mediated crystallization of the membrane protein bacterial cytochrome c oxidase. *Nat. Struct. Biol.* **2**, 842–846 (1995).
41. Sambrook, J., Fritsch, E.F. & Maniatis, T. *Molecular cloning: A laboratory manual*, edn. 2 (Cold Spring Harbor Press, New York, 1989).
42. Myska, D.G. & Morton, T.A. CLAMP: a biosensor kinetic data analysis program. *Trends Biochem. Sci.* **23**, 149–150 (1998).
43. Collaborative Computational Project, Number 4. The CCP4 Suite: Programs for Protein Crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
44. Navaza, J. *Acta Crystallogr. A* **50**, 157–163 (1994).
45. Brünger, A.T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
46. Jones, T.A., Zou, J.Y., Cowan, S.W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
47. McDonald, I.K. & Thornton, J.M. Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* **238**, 777–793 (1994).
48. Wallace, A.C., Laskowski, R.A. & Thornton, J.M. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.* **8**, 127–134 (1995).
49. Koradi, R., Billeter, M. & Wüthrich, K. MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**, 51–55 (1996).
50. Nicholls, A., Sharp, K.A. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins: Struct. Funct. Genet.* **11**, 281–296 (1991).