Protein crystallization: an overview

Zygmunt Derewenda

Department of Mol. Physiology and Biol. Physics; University of Virginia

ISFI

PSI Protein Structure Initiative

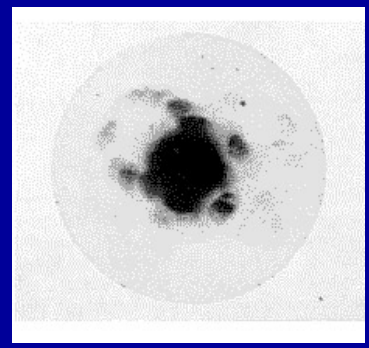
UNIVERSITY OF VIRGINIA HEALTH SYSTEM

http://techcenter.mbi.ucla.edu/

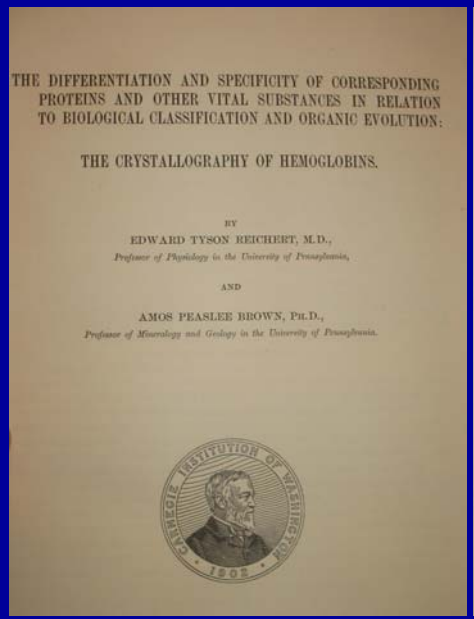
Macromolecular crystallography rests on the discovery of X-ray diffraction by inorganic crystals (something that does not occur in nature)…..

Max von Laue

Nobel Prize in Physics 1914



… and the fact that macromolecules can be crystallized in pure form (something that occurs extremely rarely in nature)…..



THE DIFFERENTIATION AND SPECIFICITY OF CORRESPONDING PROTEINS AND OTHER VITAL SUBSTANCES IN RELATION TO BIOLOGICAL CLASSIFICATION AND ORGANIC EVOLUTION:

THE CRYSTALLOGRAPHY OF HEMOGLOBINS.

BY

EDWARD TYSON REICHERT, M.D.,

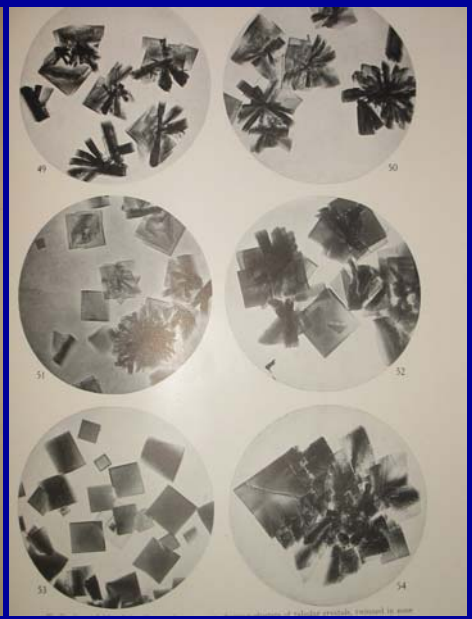
Professor of Physiology in the University of Pennsylvania,

AND

AMOS PEASLEE BROWN, PH.D.,

Professor of Mineralogy and Geology in the University of Pennsylvania.

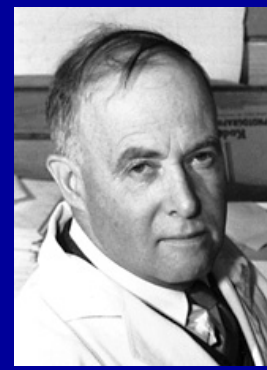
CARNEGIE INSTITUTION OF WASHINGTON 1909



Protein crystals have been know since the 19th century…

John Northrop

Wendel Stanley



James Sumner

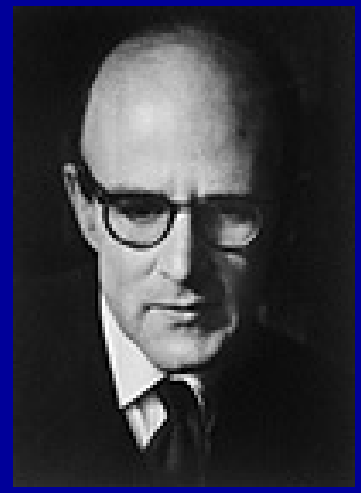
Nobel Prize in Chemistry in 1946

But it was not until 1920's that it became clear that these crystals contain pure macromolecular entities..

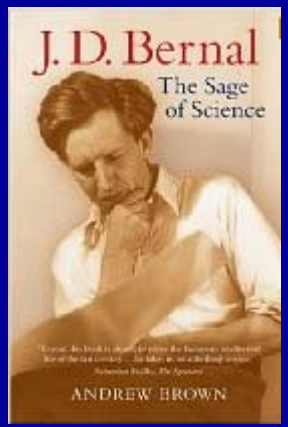
It was not until the historic work of the pioneers of protein crystallography, that the physical methods of diffraction were used on protein crystals

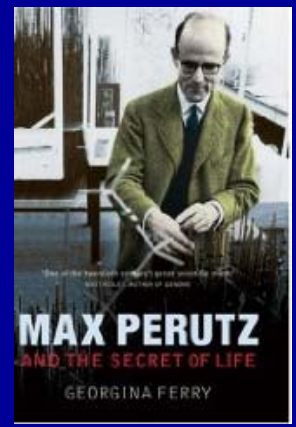
John Desmond Bernal (1901-1971)

Dorothy Crawfoot Hodgkin (1910-1995) Nobel Prize in Chemistry in 1964.



Max Ferdinand Perutz (1914-2002) Nobel Prize in Chemistry in 1962.





Historically, two alternative approaches to macromolecular crystallization were developed:

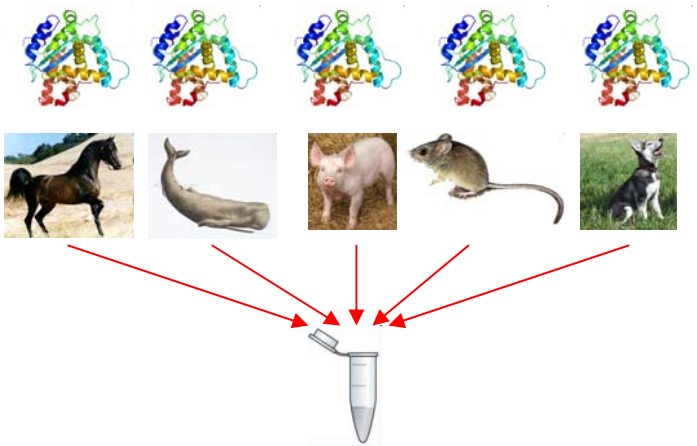
Screening homologous proteins from various sources using a standard purification and crystallization procedure……

… or screening one target protein against a large set of precipitants, buffers, etc.

NATURE November 20, 1954 Vol. 174

THE SPECIES SPECIFICITY OF MYOGLOBIN

By Dr. J. C. KENDREW and Dr. R. G. PARRISH*
Medical Research Council Unit for the Study of the Molecular Structure of Biological Systems, Cavendish Laboratory, Cambridge
AND
Prof. J. R. MARRACK and Dr. E. S. ORLANS†
Department of Pathology, University of Cambridge



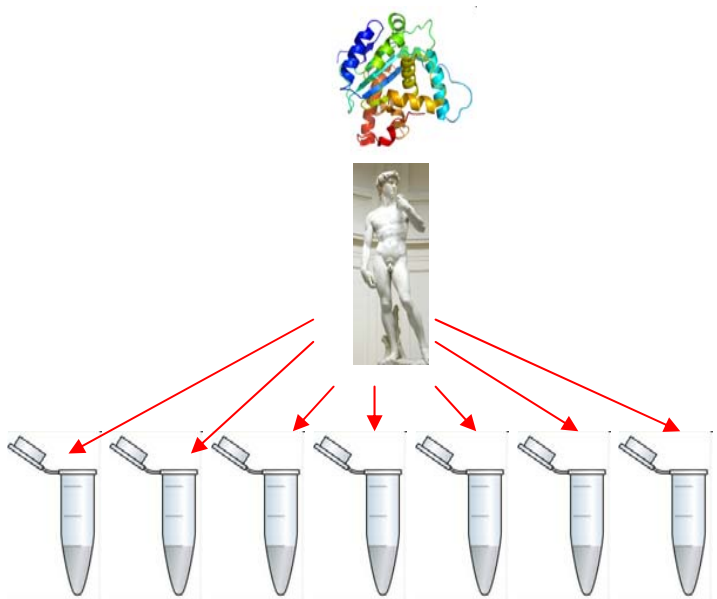
Type	Principal source	Symmetry	Space group	Mol./cell	Cell dimensions (wet) a	b	c	β
A	Sperm whale* (from ammon. sulph.)	Monoclinic	$P2_1$	2	64·6 A.	31·1 A.	34·8 A.	105·5°
B	Sperm whale (from phosphate)	Orthorhombic	$P2_12_12_1$	4	48·9	40·2	79·3	
C i	Horse†	Monoclinic	$P2_1$	2	57·3	30·8	57·0	112°
ii	Seal‡	Monoclinic	$P2_1$	2	57·7	29·6	57·1	112°
iii	Seal§	Monoclinic face-cent.	$A2$	4	57·7	29·6	106·2	102°
	Horse	Monoclinic p-orthorh.	$P2_1$	4	57·0	30·8	106·0	91°
D	Blue whale	Orthorhombic	$P2_12_12_1$	4	33·9	60·4	76·6	
E	Blue whale	Orthorhombic	$P2_12_12_1$	12–16	144·5	37·7	107·5	
F	Finback whale	Orthorhombic	$P2_12_12$	4	97·4	39·8	42·5	
G	Gentoo penguin	Face-cent. orthorhombic	$C222_1$	8	48·3	80·2	78·5	
H	Gentoo penguin (imidazole deriv., phosphate)	Orthorhombic	$P2_12_12_1$	4	106·4	39·1	45·3	
I	Gentoo penguin (imidazole deriv., ammon. sulph.)	Orthorhombic	$P2_12_12_1$ (?)	4	94·5	38·3	43·5	
J	Carp	Orthorhombic	$P2_12_12_1$	4	55·0	46·5	51·7	

The Journal of Biological Chemistry
Vol. 254, No. 23, Issue of December 10, pp. 12219–12223, 1979
Printed in U.S.A.

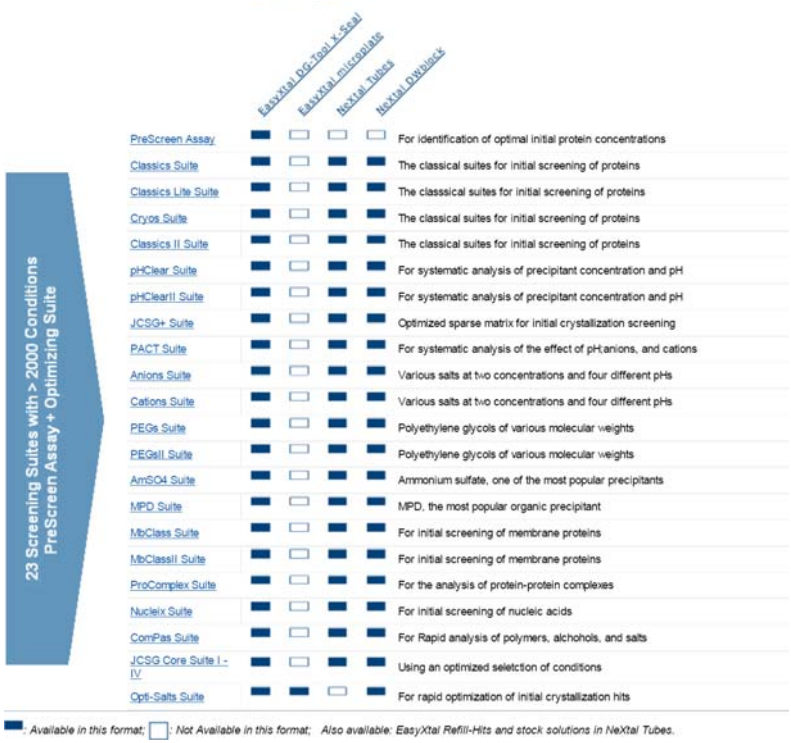
Protein Crystallization Using Incomplete Factorial Experiments*

(Received for publication, June 8, 1979)

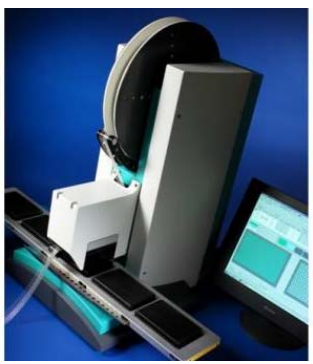
Charles W. Carter, Jr.‡ and Charles W. Carter§
From the ‡Department of Biochemistry 231H, University of North Carolina, Chapel Hill, NC 27514

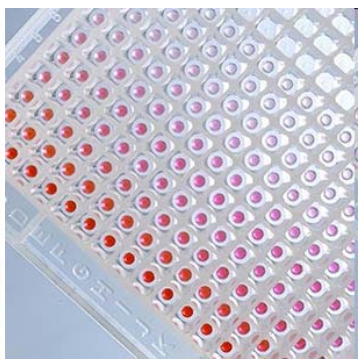


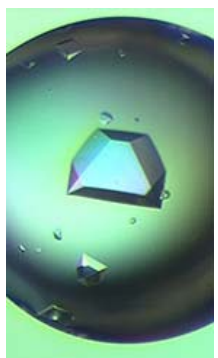
A large number of screens have been developed and are available commercially from different companies (e.g. Qiagen, below); the screens are most efficiently used with automated liquid dispensers and all follow two main principles:



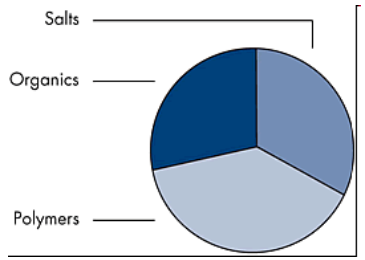
Other screen manufacturers include: Hampton Research, DeCode Genetics, Molecular Dimensions, NEXTAL, Jena Bioscience)



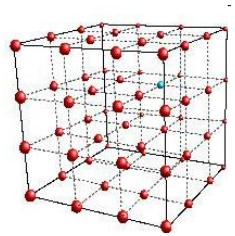




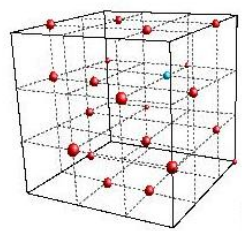
1. Salts, organic solvents and polymers (e.g. PEG) are used as precipitants and additives



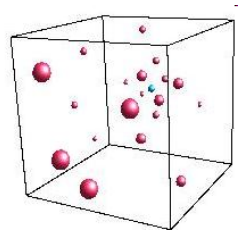
2. Screening is carried out either on full grid, incomplete factorial, or sparse matrix basis:



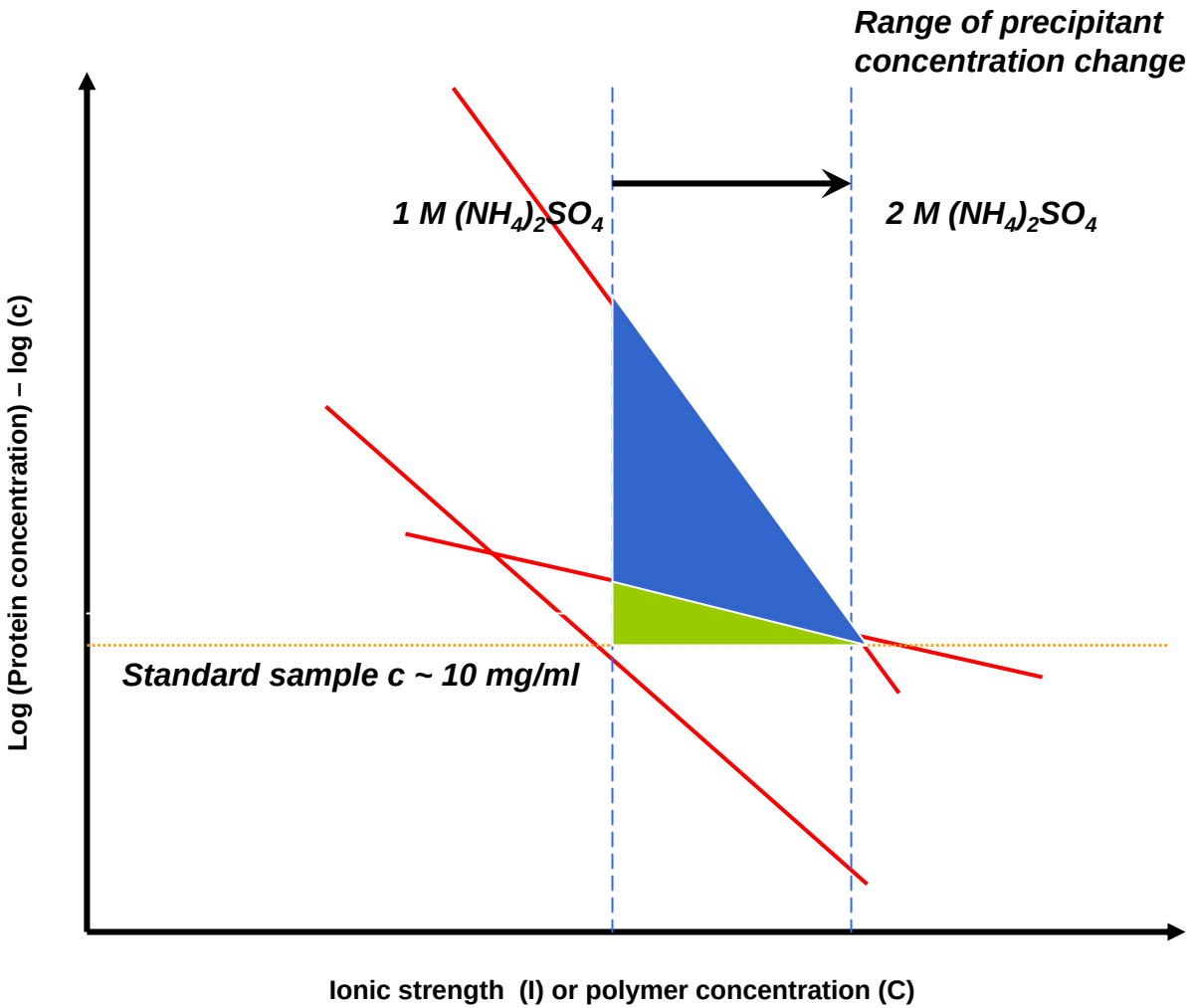
Full grid

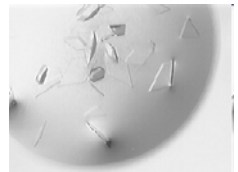


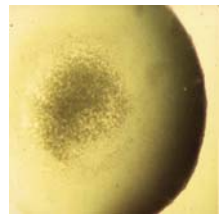
Incomplete factorial

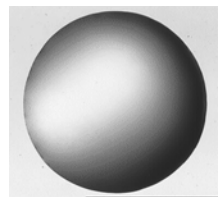


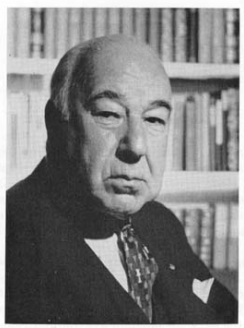
Sparse matrix: this approach relies on some prior knowledge







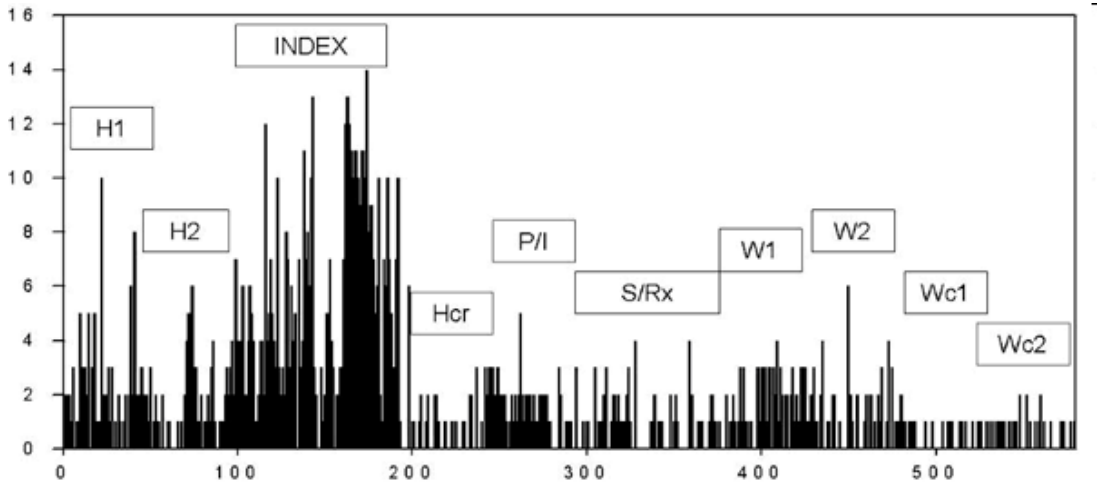




***E.J. Cohn's* equation: S, solubility; B, theoretical solubility in pure water; K, salt dependent constant; I, ionic strength of the salt**

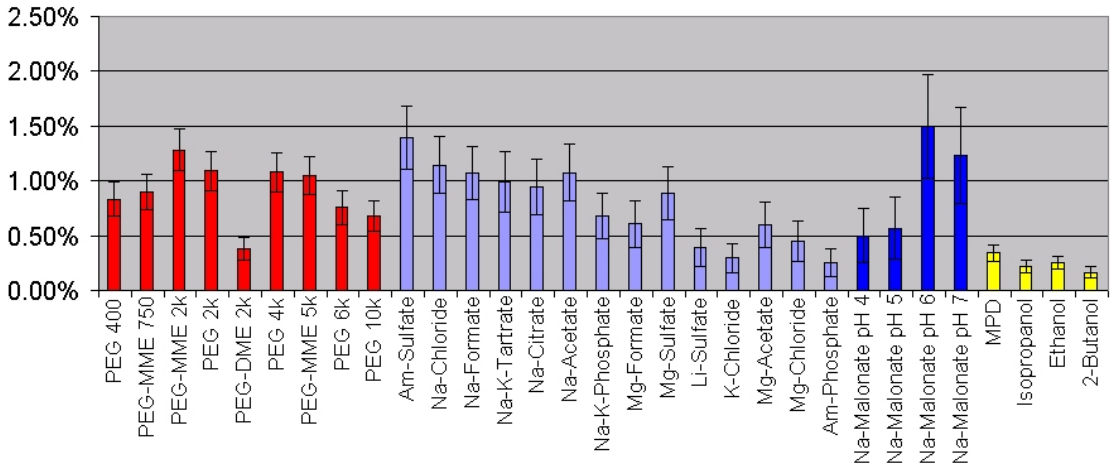
$$\log S = B - KI$$

Not all conditions in used screens are equally effective in yielding crystals. According to MCSG website ,out of 580 conditions selected from several screens, only ~300 produce crystals. This is the basis for the formulation of next generation sparse matrix screens (JCSG+, MCSG).



Data from Midwest Center for Structural Genomics

Also, various precipitants appear to show varying different intrinsic propensity for protein precipitation…



Data from: www.innovadyne.com/appnote_xtal_LLNL.html

Success rate of crystallization may be dramatically increased by the use of alternative reservoirs and additives, rather that expanding the range of conditions

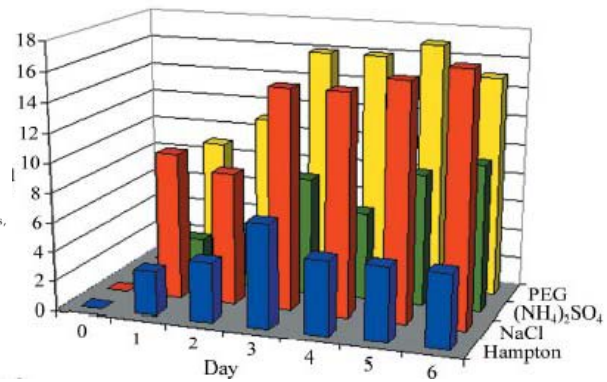
Expanding screening space through the use of alternative reservoirs in vapor-diffusion experiments

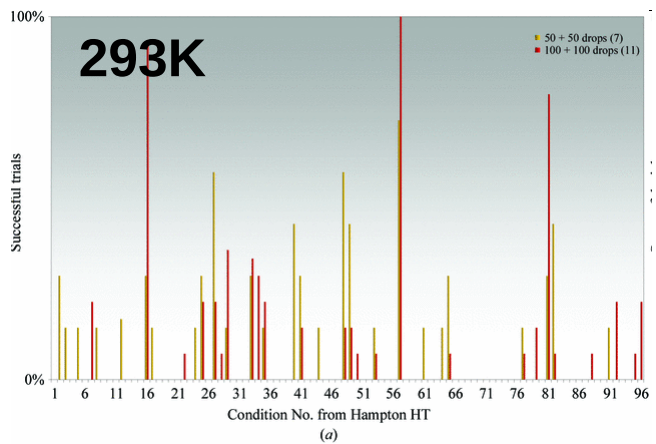
Setting up vapor-diffusion crystallization experiments against four different reservoir solutions showed that the reservoir solution may have a profound effect on the outcome of a crystallization experiment. This suggests that a facile way to increase crystallization space through screening is not to add more crystallization conditions to the process, but to set up the same conditions over different reservoirs.

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Janet Newman

In Stilla Consulting, 736 Arden Drive, Encinitas, CA 92024, USA





ELSEVIER

Available online at www.sciencedirect.com
ScienceDirect
Journal of Structural Biology 156 (2006) 387–406

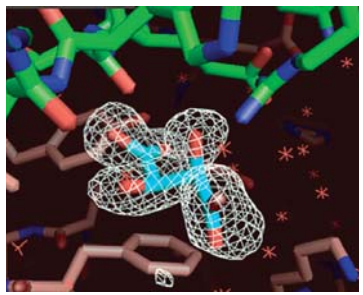
Journal of Structural Biology
www.elsevier.com/locate/yjsbi

Searching for silver bullets: An alternative strategy for crystallizing macromolecules

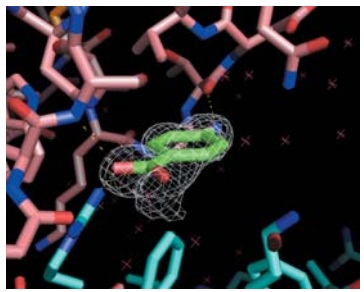
Alexander McPherson [a,*], Bob Cudney [b]

[a] University of California, Irvine, Department of Molecular Biology and Biochemistry, Room 560 Steinhaus Hall, Irvine, CA 92697-3900, USA
[b] Hampton Research, 34 Journey, Aliso Viejo, CA 92656-3317, USA

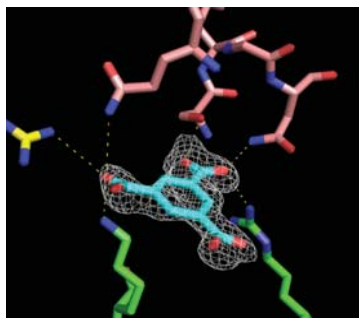
Received 24 February 2006; received in revised form 12 June 2006; accepted 14 June 2006
Available online 11 October 2006



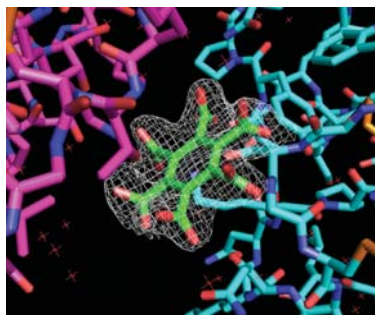
tartrate



p -aminobenzoic acid

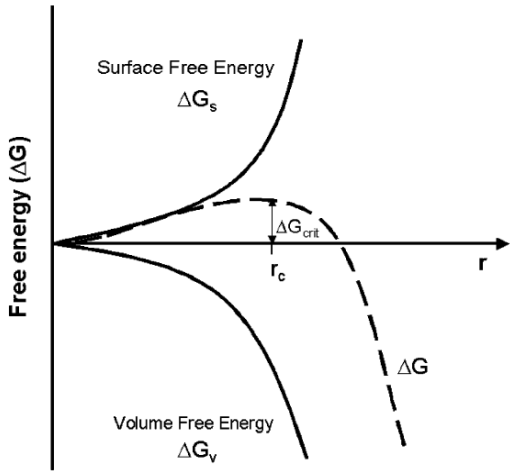


trimesic acid



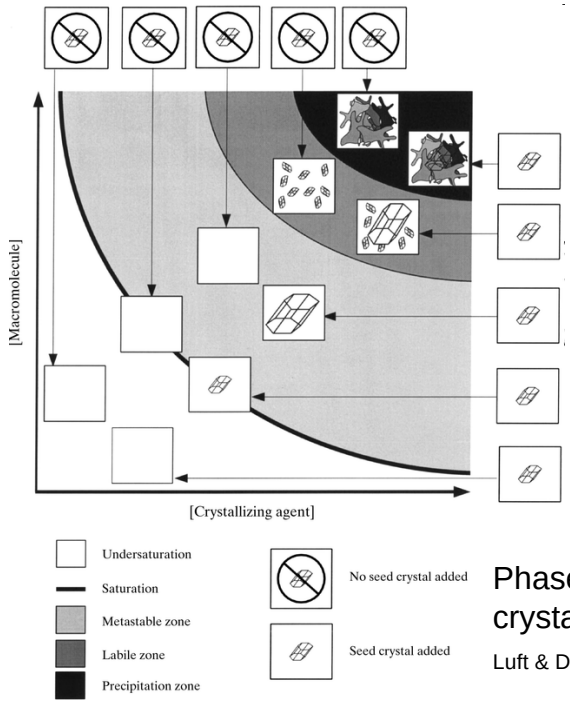
mellitic acid

A macroscopic view of crystallization helps to understand the crystallization, but only when it actually takes place – it does not help us if our target fails to crystallize as it provides few useful guidelines



Thermodynamics of nucleation according to Gibbs.

Erdemir et al (2009) Accounts Chem Res *in press*



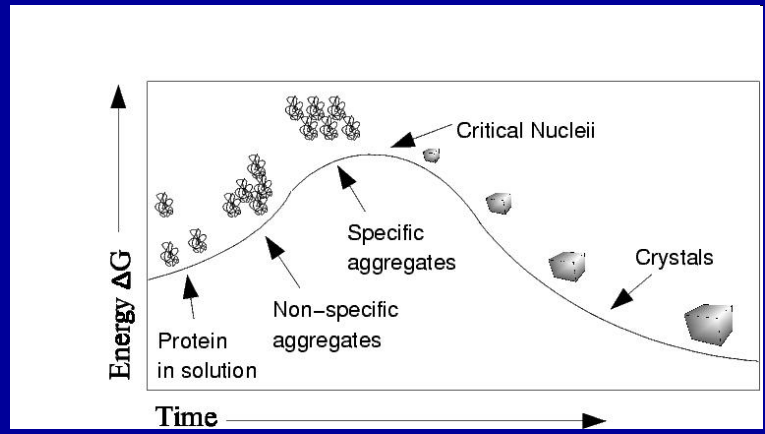
Phase diagrams for nucleation and crystallization

Luft & DeTitta (1999) Acta Cryst D55:988

A microscopic view of crystallization is required to better understand the molecular basis of protein's resistance to crystallization…

1. Nucleation of proteins is initiated from ~200-1000 % supersaturated solution in the form of clusters which undergo self reorganization to form ordered nuclei.

2. Transfer of molecules from solution to crystal is driven by small (-10 to -100 kJ mol^{-1} ; or 25 kcal) free Gibbs energy change:



$$\Delta G_{cryst} = \Delta H_{protein} - T\Delta S_{protein} - T\Delta S_{solvent}$$

This term depends on the bonds formed at the crystal contact regions, ***and is typically small***

This term may be positive and determining if enough water molecules become released from the protein's surface upon crystallization

All entropic phenomena associated with the protein during crystallization are unfavorable:

- **loss of degrees of freedom due to incorporation of molecules into the lattice which is an unavoidable entropic cost of crystallization; 30 – 100 kJ mol^{-1} at room T**
- **loss of entropy due to ordering of otherwise flexible loops, chain termini or domain flexibility; the magnitude will vary and is impossible to estimate**
- **the structure of the crystal contacts and in particular the extend of entropy loss by side chains trapped between the molecules.**

What properties confer crystallizability on a protein or protein complex?

- **The protein must have a surface that confers adequate solubility to reach supersaturation levels required for nucleation;**

- **The surface must contain patches with structured water solvent, allowing for the ordering of nascent nuclei by mediating thermodynamically viable intermolecular contacts;**

- **There should be few, if any, unstructured elements that elevate the entropic cost of crystallization, such as intrinsically disordered N- and C-termini, long partly or wholly disordered loops, or flexible carbohydrate moieties due to posttranslational modifications.**

- **Other properties, such as the Gravy index and pI may be indirectly, but positively correlated with crystallizability**

The microscopic view of crystallization permits a rational attempt to predict crystallizability from sequence features…

BIOINFORMATICS **APPLICATIONS NOTE** *Vol. 23 no. 24 2007, pages 3403–3405*
doi:10.1093/bioinformatics/btm477

Structural bioinformatics

XtalPred: a web server for prediction of protein crystallizability

Lukasz Slabinski[1,2], Lukasz Jaroszewski[1], Leszek Rychlewski[2], Ian A. Wilson[1], Scott A. Lesley[1] and Adam Godzik[1,*]

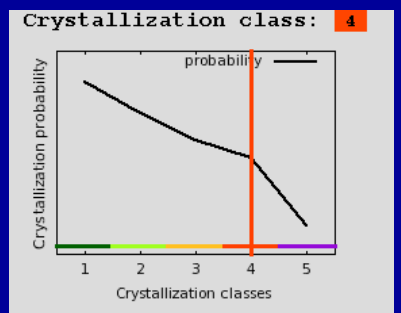
[1]Joint Center for Structural Genomics, La Jolla, CA 92037, USA and [2]BioInfoBank Institute, ul. Limanowskiego 24 A, 60-744 Poznan, Poland

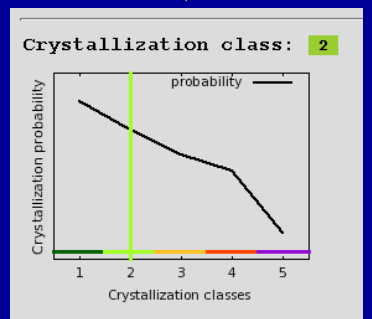
Example: human RhoGDI

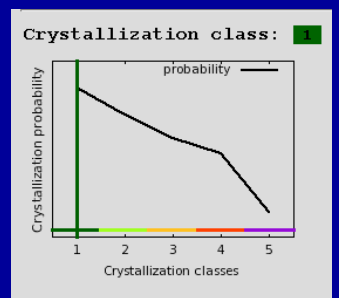
Functionally redundant

Disordered in the absence of Rho, this fragment is responsible for nucleotide exchange inhibition

Folded in solution, the C-terminal domain binds to Rho, sequesters the prenyl group, does not inhibit nucleotide exchange





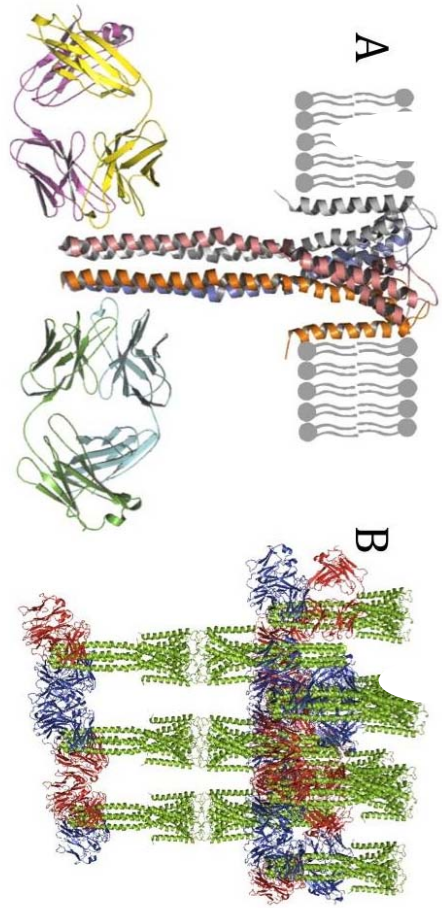


When a protein target proves recalcitrant to crystallization, we must resort to either homologue screening or the following protein engineering methods:*

- Improving solubility through:
  - Use of hybridoma or synthetic antibodies
  - Use of ‘in-line’ or insert fusion partners
  - Direct surface engineering
- Increasing stability
- Reducing the entropic barrier of crystallization through:
  - Construct optimization to reduce flexible termini or loops
  - Elimination of post-translational modifications
  - Surface entropy reduction

)A review on this subject will appear in the new Edition of the International Tables and as a Feature Article in Acta Cryst D

- Improving solubility through:
  - ***Use of hybridoma or synthetic antibodies***
  - Use of N(C)-terminal or insertion fusion partners
  - Direct surface engineering
- Increasing stability
- Reducing the entropic barrier of crystallization through:
  - Construct optimization to reduce flexible termini or loops
  - Elimination of post-translational modifications
  - Surface entropy reduction



PNAS

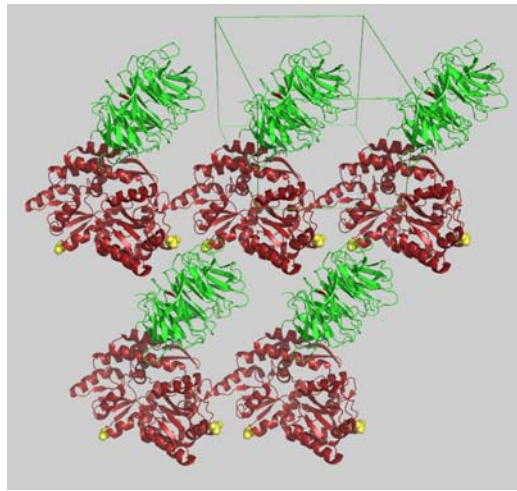
Crystal structure of full-length KcsA in its closed conformation

Serdar Uysal[a,b], Valeria Vásquez[a,b], Valentina Tereshko[a,b], Kaori Esaki[a], Frederic A. Fellouse[c], Sachdev S. Sidhu[c], Shohei Koide[a], Eduardo Perozo[a,b,1], and Anthony Kossiakoff[a,b,1]

[a]Department of Biochemistry and Molecular Biology, and [b]Institute for Biophysical Dynamics, University of Chicago, Chicago, IL 60637; and [c]Department of Protein Engineering, Genentech Inc., 1 DNA Way, South San Francisco, CA 94080

Edited by John Kuriyan, University of California, Berkeley, CA, and approved February 18, 2009 (received for review October 29, 2008)

- Improving solubility through:
  - Use of hybridoma or synthetic antibodies
  - ***Use of N(C) terminal or insertion fusion partners***
  - Direct surface engineering
- Increasing stability
- Reducing the entropic barrier of crystallization through:
  - Construct optimization to reduce flexible termini or loops
  - Elimination of post-translational modifications
  - Surface entropy reduction

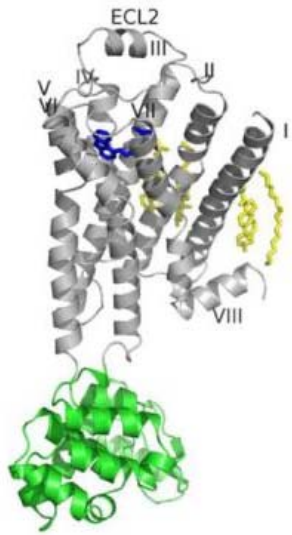


Structure of a signal transduction regulator, RACK1, from *Arabidopsis thaliana*

HEMAYET ULLAH,[1] ERICA LOUISE SCAPPINI,[2] ANDREA FLORENCE MOON,[3] LATANYA VERONICA WILLIAMS,[1,4] DAVID LEE ARMSTRONG,[2] AND LARS CHRISTIAN PEDERSEN[3]

[1]Department of Biology, Howard University, Washington, DC 20059, USA
[2]Laboratory of Neurobiology, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, North Carolina 27709, USA
[3]Laboratory of Structural Biology, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, North Carolina 27709, USA

(Received February 26, 2008; Final Revision June 20, 2008; Accepted June 25, 2008)



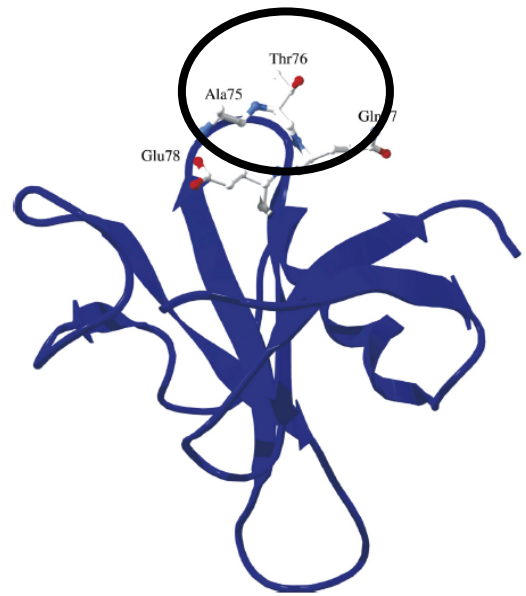
High-Resolution Crystal Structure of an Engineered Human β_2-Adrenergic G Protein–Coupled Receptor

Vadim Cherezov,[1]* Daniel M. Rosenbaum,[2]* Michael A. Hanson,[1] Søren G. F. Rasmussen,[2] Foon Sun Thian,[2] Tong Sun Kobilka,[2] Hee-Jung Choi,[2,3] Peter Kuhn,[4] William I. Weis,[2,3] Brian K. Kobilka,[2]† Raymond C. Stevens[1]†

- Improving solubility through:
  - Use of hybridoma or synthetic antibodies
  - Use of ‘in-line’ or insert fusion partners
  - ***Direct surface engineering*** ➡
- Increasing stability
- Reducing the entropic barrier of crystallization through:
  - Construct optimization to reduce flexible termini or loops
  - Elimination of post-translational modifications
  - Surface entropy reduction

What mutations should be made?

Amino acid at position 76	Solubility (mg/ml)[a]
Asp	43
Arg	42
Glu	42
Ser	39
Lys	31
Gly	27
Ala	27
His	24
Asn	21
Thr	20
Gln	20
Pro	15
Cys	12
Met	11
Val	10
Leu	9.3
Ile	8.2
Tyr	5.6
Phe	4.4
Trp	3.6



[a] The error in these measurements is ±10%.

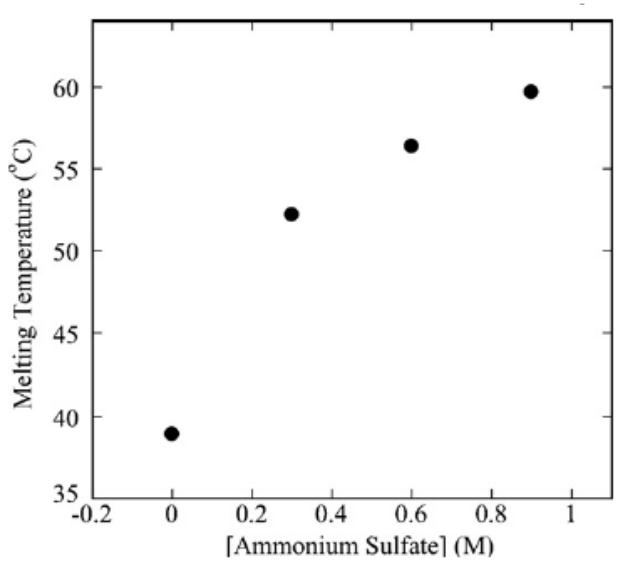
doi:10.1016/j.jmb.2006.10.026 J. Mol. Biol. (2007) 366, 449–460
JMB Available online at www.sciencedirect.com ScienceDirect ELSEVIER

Amino Acid Contribution to Protein Solubility: Asp, Glu, and Ser Contribute more Favorably than the other Hydrophilic Amino Acids in RNase Sa

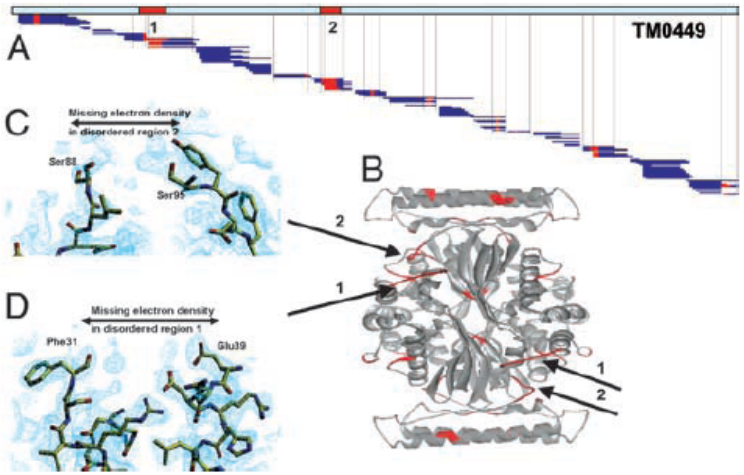
Saul R. Trevino[1], J. Martin Scholtz[1,2]* and C. Nick Pace[1,2]*

- Improving solubility through:
  - Use of hybridoma or synthetic antibodies
  - Use of ‘in-line’ or insert fusion partners
  - Direct surface engineering
- ***Increasing stability*** →
- Reducing the entropic barrier of crystallization through:
  - Construct optimization to reduce flexible termini or loops
  - Elimination of post-translational modifications
  - Surface entropy reduction

Protein	ΔH_m [a]	ΔS_m [b]	T_m [c]	ΔT_m [d]	$\Delta\Delta G^{oe}$
Wild-type	91	284	47.8		
T76P	95	293	51.4	3.6	1.0
T76Y	97	299	50.8	3.0	0.9
T76F	95	293	50.6	2.8	0.8
T76W	93	288	50.1	2.3	0.7
T76A	95	295	49.4	1.6	0.5
T76H	93	288	49.4	1.6	0.5
T76E	93	289	49.1	1.3	0.4
T76D	93	289	49.1	1.3	0.4
T76K	93	289	49.0	1.2	0.3
T76L	94	292	48.8	1.0	0.3
T76Q	94	292	48.6	0.8	0.2
T76S	92	286	48.6	0.8	0.2
T76R	94	292	48.5	0.7	0.2
T76M	91	283	48.4	0.6	0.2
T76N	89	277	48.4	0.6	0.2
T76V	93	289	48.2	0.4	0.1
T76I	92	286	48.0	0.2	0.1
T76C	64	199	47.7	−0.1	0.0
T76G	88	275	46.9	−0.9	−0.3



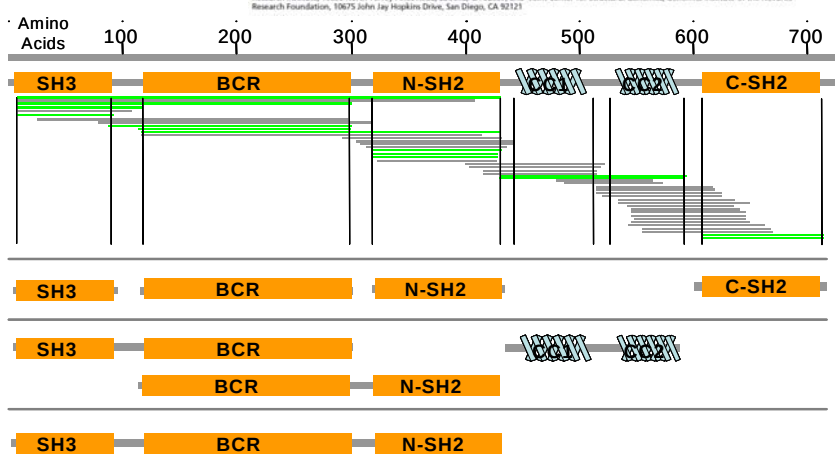
- Improving solubility through:
  - Use of hybridoma or synthetic antibodies
  - Use of ‘in-line’ or insert fusion partners
  - Direct surface engineering
- Increasing stability
- Reducing the entropic barrier of crystallization through:
  - ***Construct optimization to reduce flexible termini or loops***
  - Elimination of post-translational modifications
  - Surface entropy reduction



Rapid refinement of crystallographic protein construct definition employing enhanced hydrogen/deuterium exchange MS

Dennis Pantazatos*, Jack S. Kim*, Heath E. Klock†, Raymond C. Stevens‡, Ian A. Wilson‡, Scott A. Lesley†, and Virgil L. Woods, Jr.*¶§

*Department of Medicine, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA, 92093; †Joint Center for Structural Genomics, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037; and ‡Joint Center for Structural Genomics, Genomics Institute of the Novartis Research Foundation, 10675 John Jay Hopkins Drive, San Diego, CA 92121



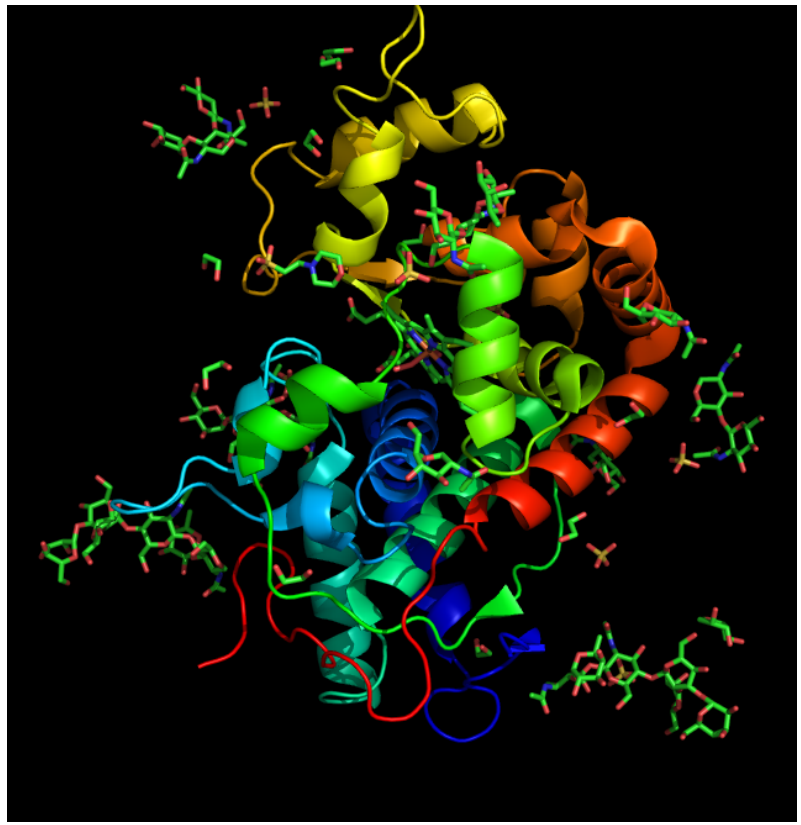
Automated, high-throughput platform for protein solubility screening using a split-GFP system

Pawel Listwan · Thomas C. Terwilliger · Geoffrey S. Waldo

- Improving solubility through:
  - Use of hybridoma or synthetic antibodies
  - Use of ‘in-line’ or insert fusion partners
  - Direct surface engineering
- Increasing stability
- Reducing the entropic barrier of crystallization through:
  - Construct optimization to reduce flexible termini or loops
  - ***Elimination of post-translational modifications*** →
  - Surface entropy reduction

N-X-S/T mutations

Not always necessary…

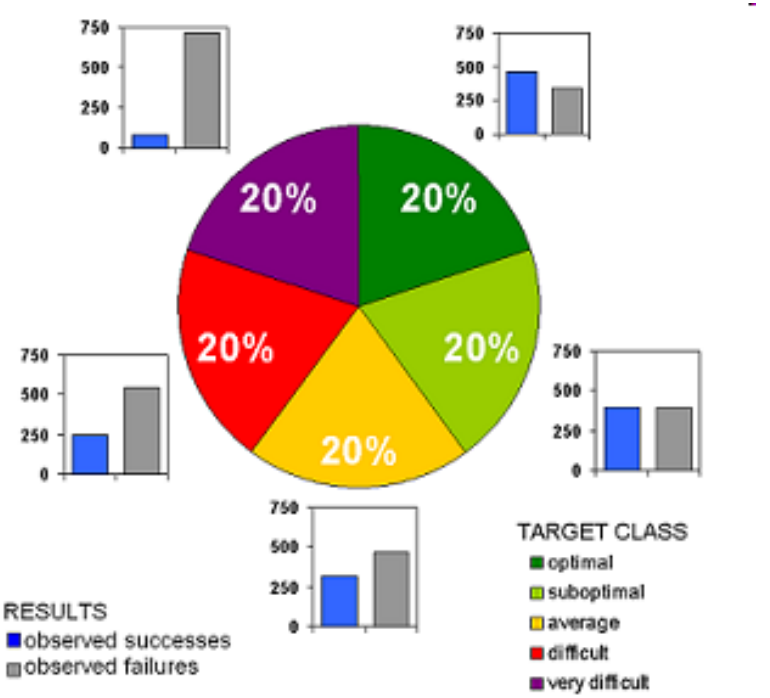


Crystal structure and statistical coupling analysis of highly glycosylated peroxidase from royal palm tree (*Roystonea regia*)

Leandra Watanabe [a], Patricia Ribeiro de Moura [a], Lucas Bleicher [a], Alessandro S. Nascimento [a], Laura S. Zamorano [b], Juan J. Calvete [c], Lìbia Sanz [c], Alicia Pérez [c], Sergey Bursakov [d], Manuel G. Roig [b], Valery L. Shnyrov [e,*], Igor Polikarpov [a,*]

[a] Instituto de Física de São Carlos, Departamento de Física e Informática, Universidade de São Paulo, Avenida Trabalhador São Carlense 400, CEP 13566-590 São Carlos, SP, Brazil
[b] Departamento de Química Física, Facultad de Química, Universidad de Salamanca, 37008 Salamanca, Spain
[c] Instituto de Biomedicina de Valencia (CSIC), 46010 Valencia, Spain
[d] REQUIMTE, Departamento de Química, Centro de Química Fina e Biotecnologia, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, 2829-516 Caparica, Portugal
[e] Departamento de Bioquímica y Biología Molecular, Facultad de Biología, Universidad de Salamanca, 37007 Salamanca, Spain

However, even the best behaving proteins seem to have at best 50/50 chance of forming crystals, while experimental data from SG pipelines shows that only ~10 of purifiable protein yield X-ray quality crystals



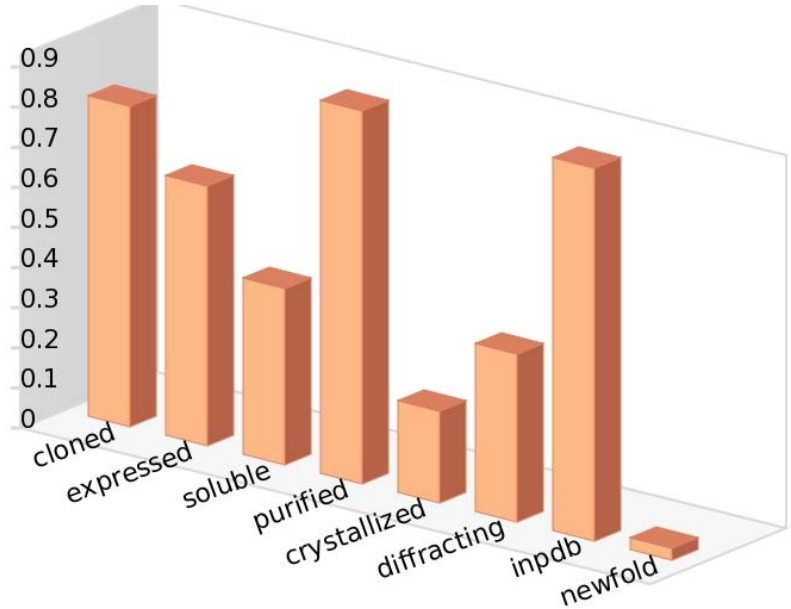
BIOINFORMATICS APPLICATIONS NOTE Vol. 23 no. 24 2007, pages 3403–3405 doi:10.1093/bioinformatics/btm477

Structural bioinformatics

XtalPred: a web server for prediction of protein crystallizability

Lukasz Slabinski[1,2], Lukasz Jaroszewski[1], Leszek Rychlewski[2], Ian A. Wilson[1], Scott A. Lesley[1] and Adam Godzik[1,*]

[1]Joint Center for Structural Genomics, La Jolla, CA 92037, USA and [2]BioInfoBank Institute, ul. Limanowskiego 24 A, 60-744 Poznan, Poland



Data from http://www.mcsg.anl.gov/

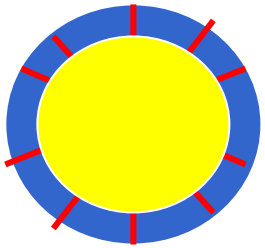
The role of entropy and polarity in intermolecular contacts in protein crystals

Marcin Cieślik and Zygmunt S. Derewenda

Acta Cryst. (2009). D**65**, 500–509

A statistical analysis of crystal contacts in 821 unambiguously monomeric proteins crystallized in 51 different space groups

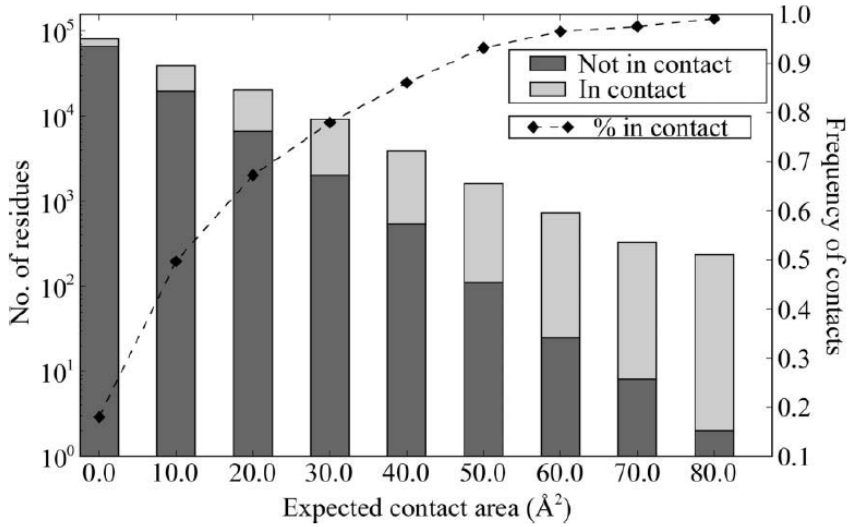
Is the crystal-contact forming propensity directly proportional to the solvent accessible surface area presented by a particular amino acid?



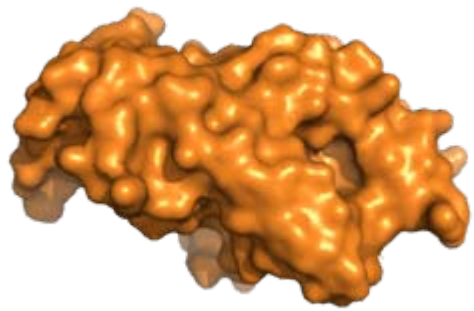
The expected contact area (ECA)

$$\mathrm{rECA} = \mathrm{rASA}\left(\frac{\Delta\mathrm{ASA}}{\mathrm{ASA}}\right)$$

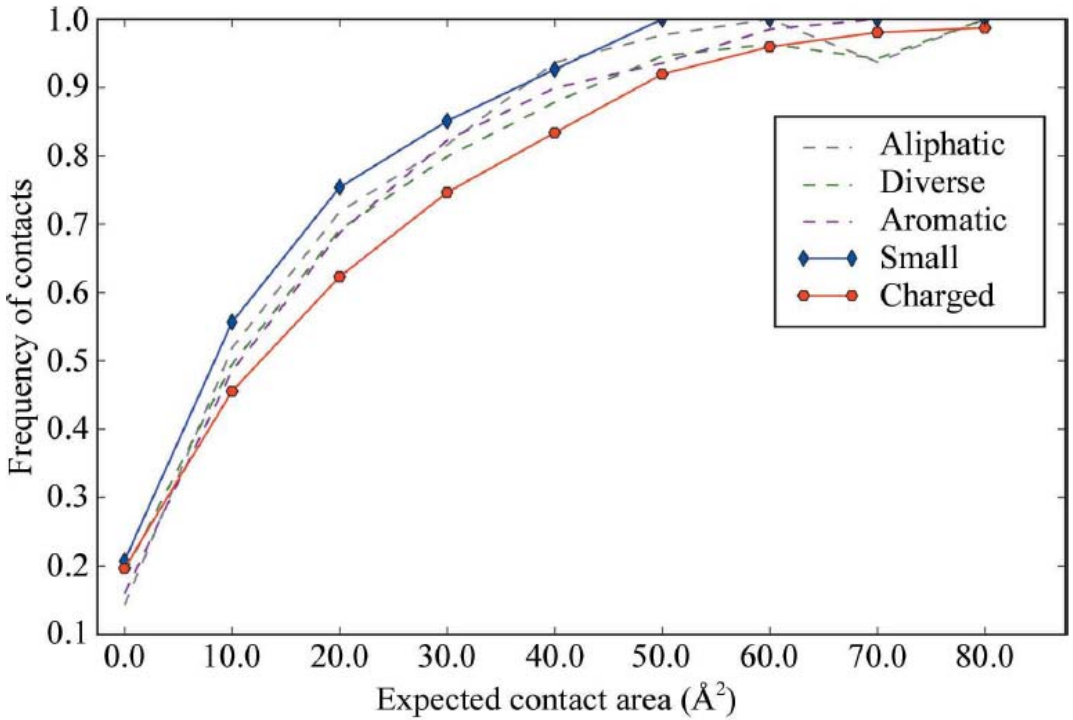
(*a*) Relative frequencies of five categories of amino acids, *i.e.* aliphatic (Val, Leu, Ile), aromatic (Trp, Phe, Tyr, His), small (Ala, Gly, Ser, Thr, Cys), charged (Lys, Arg, Glu, Asp) and other (Asn, Gln, Met, Pro), binned as a function of rECA. The relative frequency in each bin is the ratio of the number of residues of a given type to the total number of residues. (*b*) The fraction of residues involved in crystal contacts as a function of rECA plotted for the five categories as defined above.



No, the relationship is not linear. The more buried an amino acid is, the less likely it is to form a contact. This can be rationalized in terms of surface topology which seriously deviates from ideality.



Are crystal-contact forming propensities a function of physicochemical properties of amino acids?



(*a*) Relative frequencies of five categories of amino acids, *i.e.* aliphatic (Val, Leu, Ile), aromatic (Trp, Phe, Tyr, His), small (Ala, Gly, Ser, Thr, Cys), charged (Lys, Arg, Glu, Asp) and other (Asn, Gln, Met, Pro), binned as a function of rECA. The relative frequency in each bin is the ratio of the number of residues of a given type to the total number of

Given the same exposed surface, small and hydrophobic amino acids have larger propensity to form crystal contacts than charged residues.

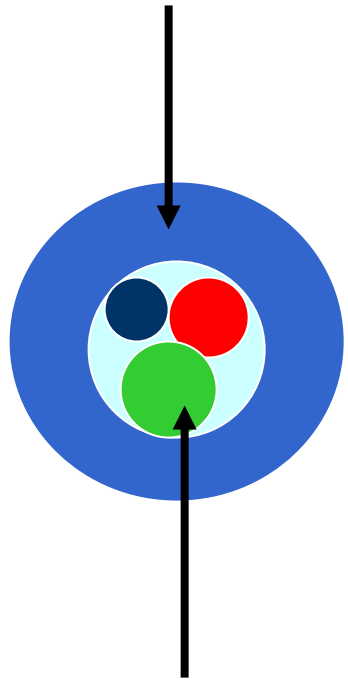
Logistic regression

$$\ln\left[\frac{Pi}{1-Pi}\right]=\alpha+\beta_{polarity}\,POLARITY_i+\beta_{sce}\,SCE_i+\beta_{charge}\,CHARGE_i+\beta_{rECA}\,rECA_i$$

Amino acid		Maximum-likelihood estimates of parameters β_{rECA}†	α‡	Contact core (%)	Contact rim (%)	Contact total (%)	Contact surface (%)	POL§	SCE¶
Gly	Asp	61 ± 0.010	−1.937 ± 0.098	4.40	4.58	4.55	4.52	11	0.00
Leu	Arg	51 ± 0.009	−2.324 ± 0.099	6.71	4.11	4.53	4.42	3	0.71
Ile	Glu	50 ± 0.011	−2.322 ± 0.125	4.12	2.26	2.56	2.41	1	0.76
Val	Ser	49 ± 0.010	−2.225 ± 0.110	4.72	2.95	3.23	3.18	4	0.43
Ala	Lys	49 ± 0.009	−2.067 ± 0.101	4.60	4.57	4.58	4.49	9	0.00
Phe	Gly	46 ± 0.012	−2.256 ± 0.136	4.55	2.43	2.77	2.35	2	0.62
Cys	Ala	44 ± 0.021	−1.941 ± 0.196	0.81	0.54	0.59	0.66	7	0.85
Tyr	His	30 ± 0.010	−2.098 ± 0.135	5.75	3.58	3.92	3.47	8	1.13
Ser	Asn	29 ± 0.008	−1.862 ± 0.106	5.51	5.53	5.53	5.52	14	1.11
Met	Thr	25 ± 0.014	−2.304 ± 0.200	1.97	1.58	1.72	1.51	5	1.46
Trp	Gln	23 ± 0.016	−2.170 ± 0.211	2.16	1.46	1.58	1.33	6	0.99
Pro	Pro	18 ± 0.008	−1.870 ± 0.125	5.24	5.49	5.45	5.22	13	0.06
Thr	Cys	15 ± 0.008	−1.819 ± 0.108	6.35	5.26	5.44	5.50	12	1.08
His	Met	05 ± 0.011	−1.900 ± 0.169	2.95	2.46	2.53	2.55	10	0.95
Asn	Val	05 ± 0.007	−1.768 ± 0.123	5.57	6.82	6.62	6.31	16	1.03
Asp	Leu	98 ± 0.006	−1.673 ± 0.107	6.18	8.07	7.77	8.35	19	0.78
Gln	Ile	94 ± 0.007	−1.711 ± 0.141	5.70	6.13	6.06	5.79	17	1.73
Arg	Tyr	86 ± 0.006	−1.684 ± 0.128	8.60	9.66	9.49	8.61	15	1.88
Glu	Phe	84 ± 0.005	−1.624 ± 0.112	7.19	10.79	10.21	10.83	18	1.46
Lys	Trp	74 ± 0.005	−1.545 ± 0.116	6.82	11.64	10.87	13.00	20	1.89

† Slope of residue expected contact area (rECA). ‡ Intercept. § Polarity scale from Trinquier & Sanejouand (1998). ¶ Side-chain entropy scale from Doig & Sternberg (1995).

Contact rim



Contact core

Molecular Dynamics Characterization of Protein Crystal Contacts in Aqueous Solutions

Giuseppe Pellicane
Dipartimento di Fisica, Università degli Studi di Messina, Contrada Papardo, 98166 Messina, Italy

Graham Smith
Henry Wellcome Laboratory for Biogerontology, Newcastle University, Newcastle upon Tyne, NE4 5PL, United Kingdom

Lev Sarkisov*
Institute for Materials and Processes, School of Engineering and Electronics, University of Edinburgh, Edinburgh, EH9 3JL, United Kingdom

(Received 2 October 2008; published 10 December 2008)

We employ nonequilibrium molecular dynamics simulation to characterize the effective interactions between lysozyme molecules involved in the formation of two hydrophobic crystal contacts. We show that the effective interactions between crystal contacts do not exceed a few kT, the range of the attractive part of the potential is less than 4 Å, and, within this range, there is a significant depletion of water density between two protein contacts. Our findings highlight the different natures of protein crystallization and protein recognition processes.

The possible role and nature of anisotropic interactions in protein crystallization has been the subject of intense investigation. Indeed, the effort of the recent decades to understand protein interactions in aqueous solution and to describe quantitatively the phase diagram by means of isotropic models has proven to be an elusive task [1–8]. Globular proteins are not perfectly spherical, and their surface is structurally and energetically heterogeneous, leading to anisotropic protein-protein interactions.

No unified picture of these anisotropic interactions has been developed to date. A popular direction of research associates anisotropic contributions in water-mediated (solvation) protein-protein interactions with hydrophobic regions (patches) on the surfaces of the proteins [9,10]. Often, a "patch-patch interaction" term is introduced in the model to incorporate these effects [2,4,6,7,11–13].

crystal contact [26]. However, our observations do not support a picture of protein crystallization as a purely stochastic process as has been also suggested by Carugo and Argos [25]. Indeed, it is quite evident that the magnitude of the observed interactions and even their shape *do* depend on the nature of the interfaces participating in the contact formation, thus leading to a slight anisotropy of protein interactions. Larger surfaces involving positively

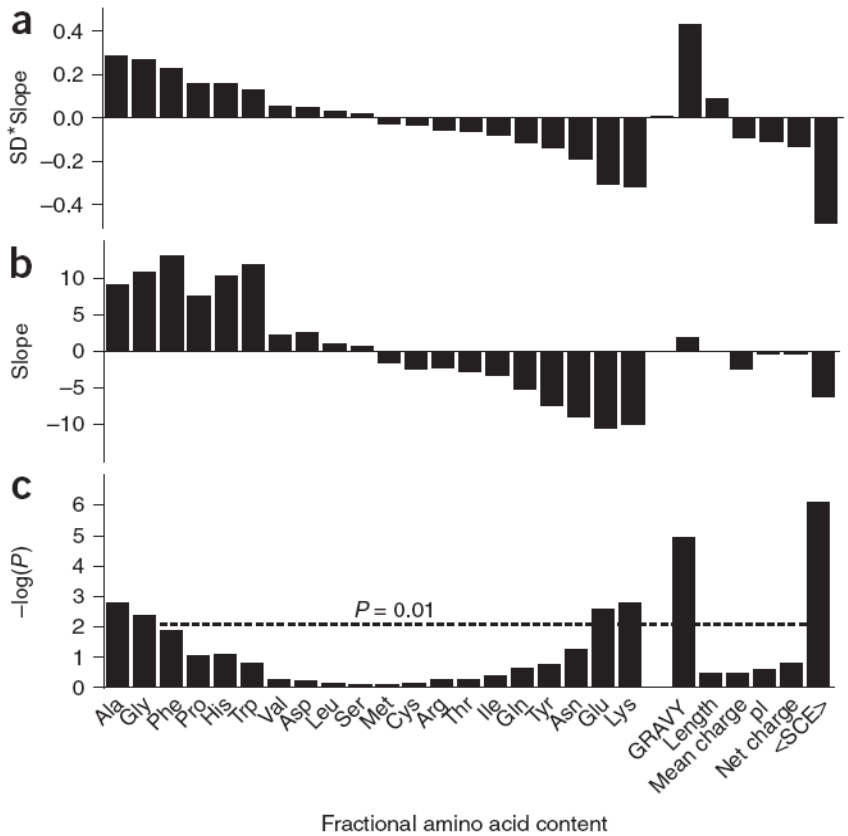
Understanding the physical properties that control protein crystallization by analysis of large-scale experimental data

A recent statistical analysis of 679 well expressed proteins, of which 157 yielded crystal structures

W Nicholson Price II[1,2], Yang Chen[1,2], Samuel K Handelman[1,2], Helen Neely[1,2], Philip Manor[1,2], Richard Karlin[1,2], Rajesh Nair[1,3], Jinfeng Liu[1,3], Michael Baran[1,4], John Everett[1,4], Saichiu N Tong[1,4], Farhad Forouhar[1,2], Swarup S Swaminathan[1,2], Thomas Acton[1,4], Rong Xiao[1,4], Joseph R Luft[1,5], Angela Lauricella[1,5], George T DeTitta[1,5], Burkhard Rost[1,3], Gaetano T Montelione[1,4,6] & John F Hunt[1,2]

NATURE BIOTECHNOLOGY VOLUME 27 NUMBER 1 JANUARY 2009

Figure 3 Correlations between sequence characteristics and success in crystal structure solution. Logistic regressions based on success in crystal structure determination (that is, PDB deposition) were performed on a dataset comprising 679 proteins from the NESG protein expression and crystallization pipeline. Variables evaluated included the fractional content of each amino acid, mean residue hydrophobicity (GRAVY[28]), chain length, mean charge (fraction arginine+lysine+asparagine+glutamatic acid), pI, mean net charge and 〈SCE〉. (**a**) Predictive value of each parameter, which is defined as the product of its logistic regression slope and the s.d. of its distribution in the dataset. (**b**) Logistic regression slope. (**c**) Negative log of logistic regression *P*-value.



"Our statistical analysis of large-scale protein crystallization results demonstrates that the mean entropy of exposed side chains and predicted backbone disorder both anti-correlate strongly and significantly with successful structure determination. Combining these results with the observation that stability is not a significant determinant of success leads to the conclusion that the dominant factor determining protein crystallization outcome is the prevalence of well-ordered surface epitopes capable of mediating stereochemically specific interprotein packing interactions".

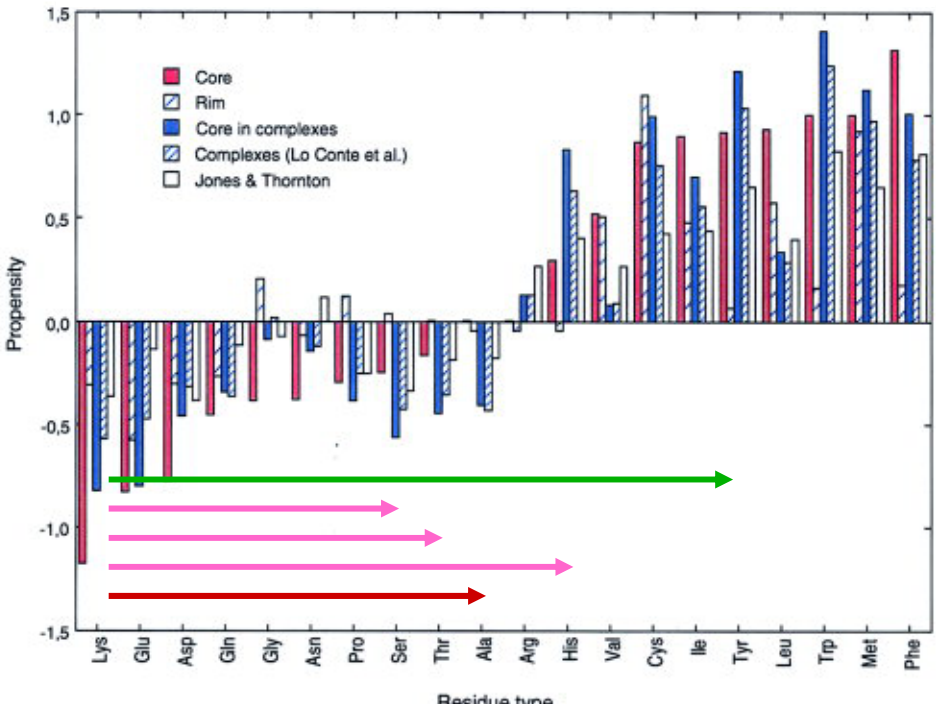
The option of last resort for poorly or non-crystallizable proteins:

- Improving solubility through:
  - Use of hybridoma or synthetic antibodies
  - Use of ‘in-line’ or insert fusion partners
  - Direct surface engineering
- Increasing stability
- Reducing the entropic barrier of crystallization through:
  - Construct optimization to reduce flexible termini or loops
  - Elimination of post-translational modifications
  - ***Surface entropy reduction***

What amino acids should favor intermolecular contacts?

Studies of protein-protein complexes and antigen-antibody complexes suggest that Ala, Tyr, Ser and His might effectively replace Lys, Glu and Gln to generate crystal contact forming epitopes on proteins' surfaces

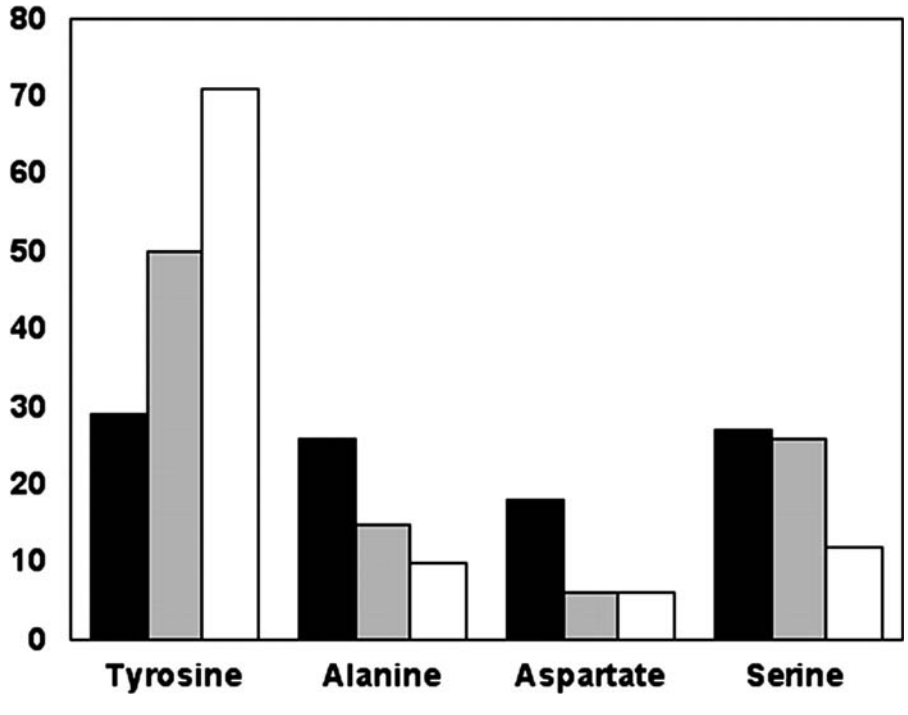
PROTEINS: Structure, Function, and Genetics 53:708–719 (2003)



Dissecting Subunit Interfaces in Homodimeric Proteins

Ranjit Prasad Bahadur,[1] Pinak Chakrabarti,[1] Francis Rodier,[2] and Joël Janin[2]*
[1]*Department of Biochemistry, Bose Institute, Calcutta, India*
[2]*Laboratoire d'Enzymologie et de Biochimie Structurales, CNRS UPR 9063, Gif-sur-Yvette, France*

PNAS | August 24, 2004 | vol. 101 | no. 34 | 12467–12472

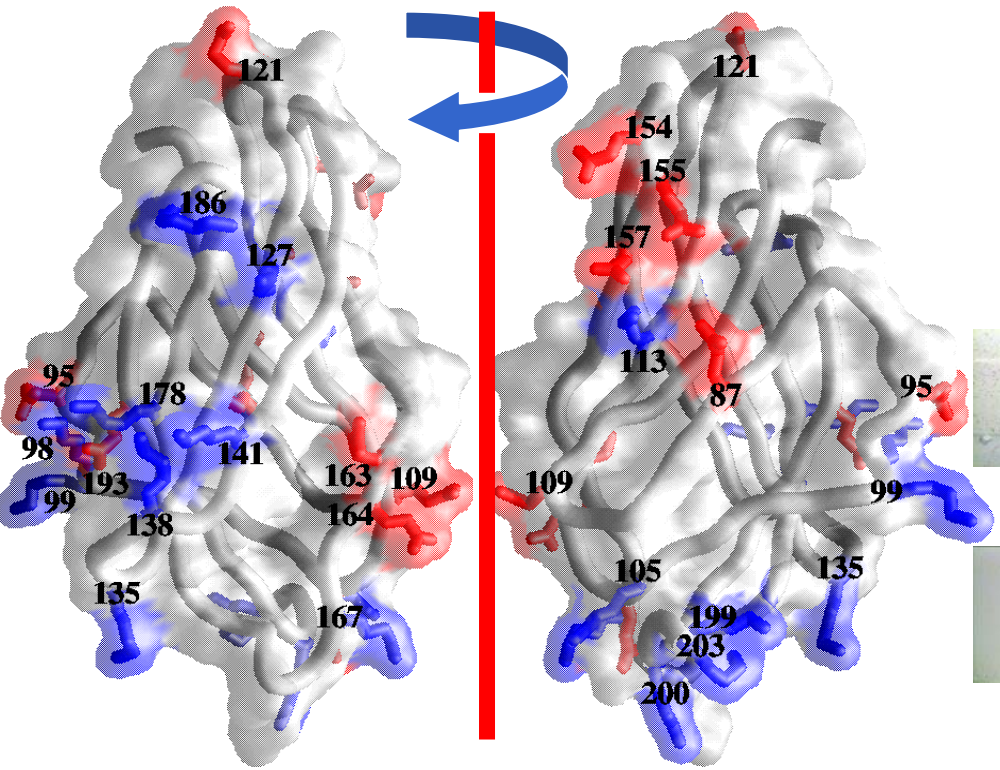


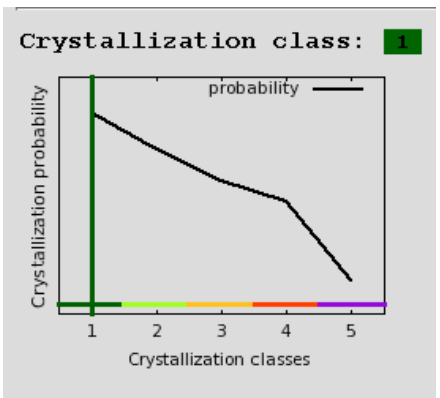
Synthetic antibodies from a four-amino-acid code: A dominant role for tyrosine in antigen recognition

Frederic A. Fellouse*, Christian Wiesmann*, and Sachdev S. Sidhu*†

*Department of Protein Engineering, Genentech Inc., 1 DNA Way, South San Francisco, CA 94080

The impact of mutations of Lys and Glu residues to Ala was tested using RhoGDI (~10% Lys and ~10% Glu content)





Longenecker, et al. (2001) Acta Crystallogr D57:679-88.
Mateja et al. (2002) Acta Crystallogr D58:1983-91

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

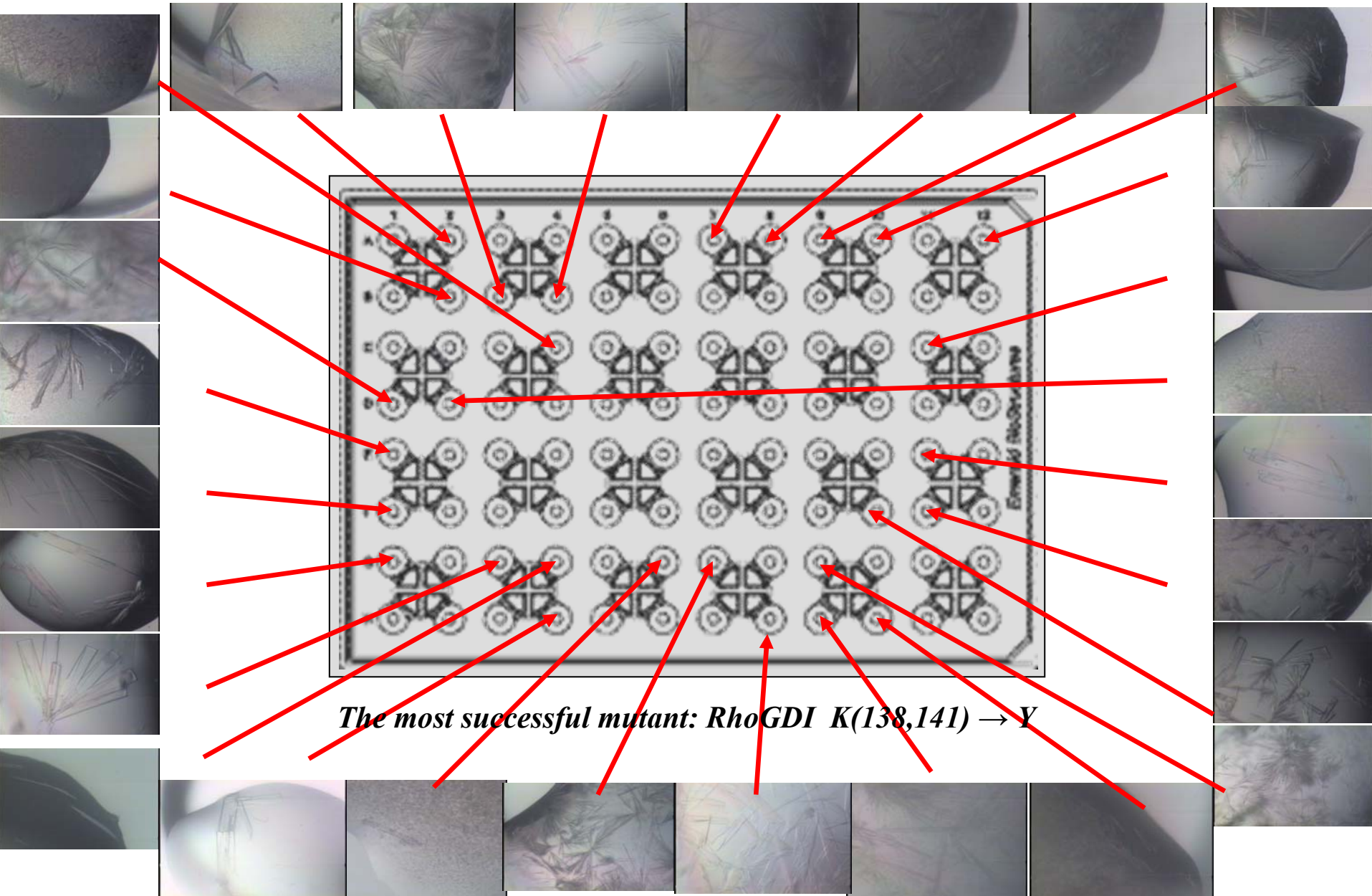
Protein crystallization by surface entropy reduction: optimization of the SER strategy

	Ala			His			Ser			Thr			Tyr			Totals		
	Standard	Salt	Unique	Standard	Salt	Unique	Standard	Salt	Unique	Standard	Salt	Unique	Standard	Salt	Unique	Standard	Unique	Salt	Unique
A	1	0	1	2	2	3	1	18	19	5	15	17	8	13	17	17	14	48	31
C	2	9	9	2	2	4	1	2	3	3	14	16	5	5	6	13	12	32	26
D	1	1	1	4	7	11	0	0	0	0	1	1	34	35	48	39	39	44	37
E	6	0	6	1	0	1	2	0	2	5	4	8	0	2	2	14	8	6	6
F	11	10	16	3	0	3	3	3	6	1	0	1	0	0	0	18	11	13	10
G	0	4	4	5	8	11	1	6	7	2	2	4	14	12	20	22	17	32	18
H	12	15	20	2	1	3	1	1	2	0	1	1	17	3	20	32	30	21	18
I	4	5	8	2	4	5	3	6	7	16	28	35	3	3	6	28	24	46	35
	37	44	65	21	24	41	12	36	46	32	65	83	81	73	119				

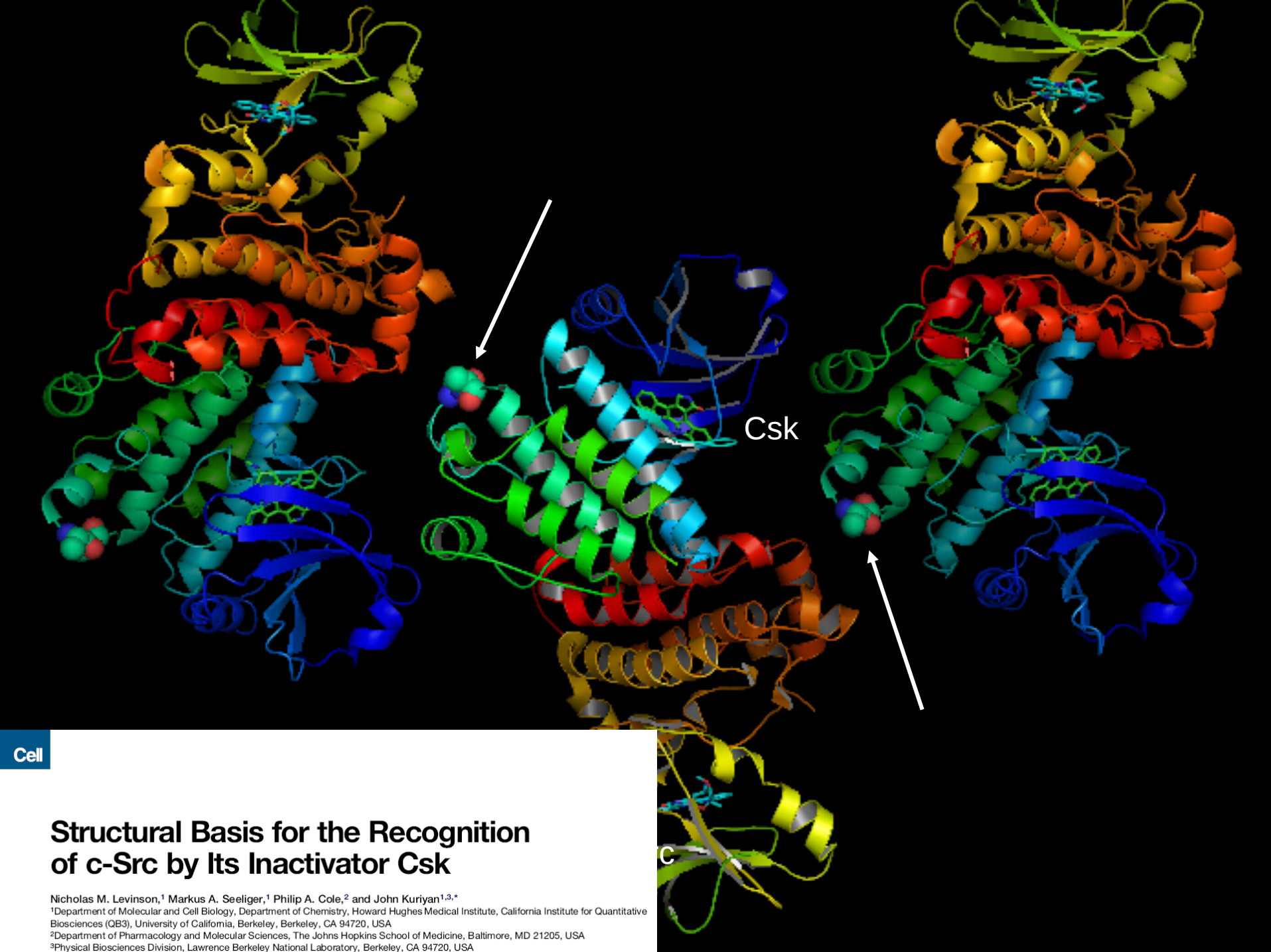
Acta Cryst. (2007). D**63**, 636–645

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Protein crystallization by surface entropy reduction: optimization of the SER strategy



The most successful mutant: RhoGDI K(138,141) → Y



Cell

Structural Basis for the Recognition of c-Src by Its Inactivator Csk

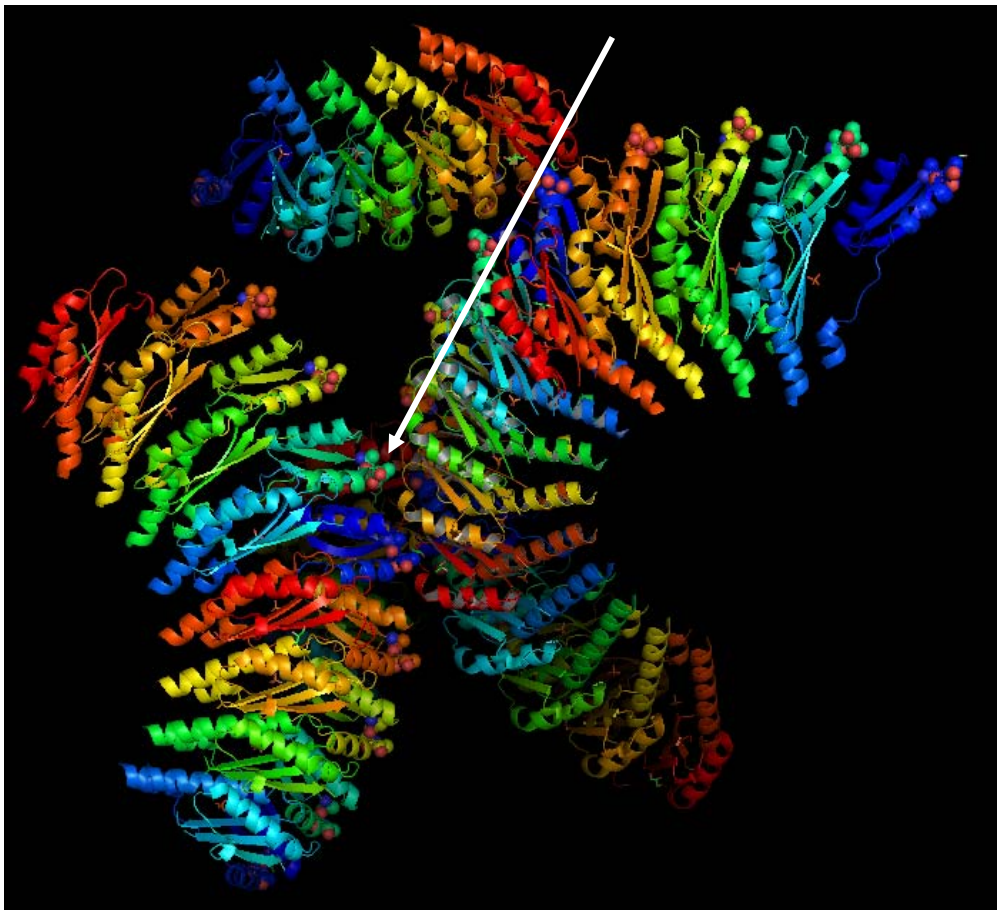
Nicholas M. Levinson,[1] Markus A. Seeliger,[1] Philip A. Cole,[2] and John Kuriyan[1,3,*]
[1]Department of Molecular and Cell Biology, Department of Chemistry, Howard Hughes Medical Institute, California Institute for Quantitative Biosciences (QB3), University of California, Berkeley, Berkeley, CA 94720, USA
[2]Department of Pharmacology and Molecular Sciences, The Johns Hopkins School of Medicine, Baltimore, MD 21205, USA
[3]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

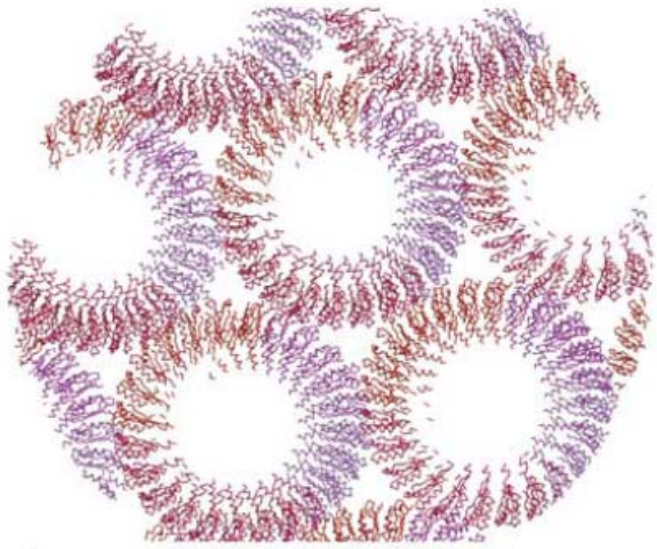
nature Vol 435|2 June 2005|doi:10.1038/nature03554

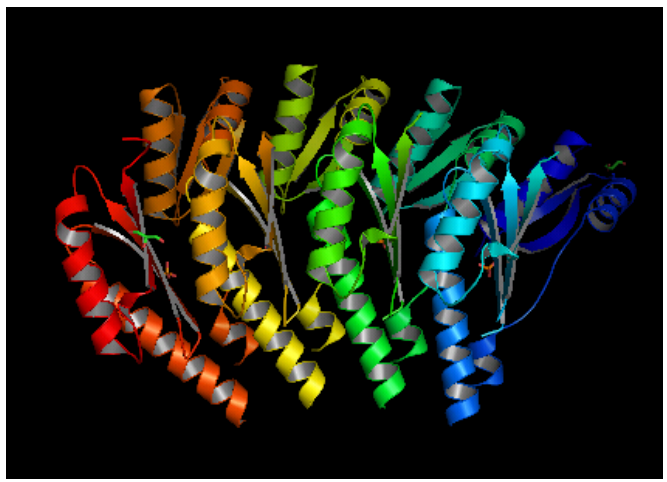
LETTERS

Structural characterization of the molecular platform for type III secretion system assembly

Calvin K. Yip[1], Tyler G. Kimbrough[2], Heather B. Felise[3], Marija Vuckovic[1], Nikhil A. Thomas[4], Richard A. Pfuetzner[1], Elizabeth A. Frey[1], B. Brett Finlay[4], Samuel I. Miller[2,3] & Natalie C. J. Strynadka[1]







doi:10.1016/j.jmb.2007.10.035 J. Mol. Biol. (2008) 375, 471–486

JMB

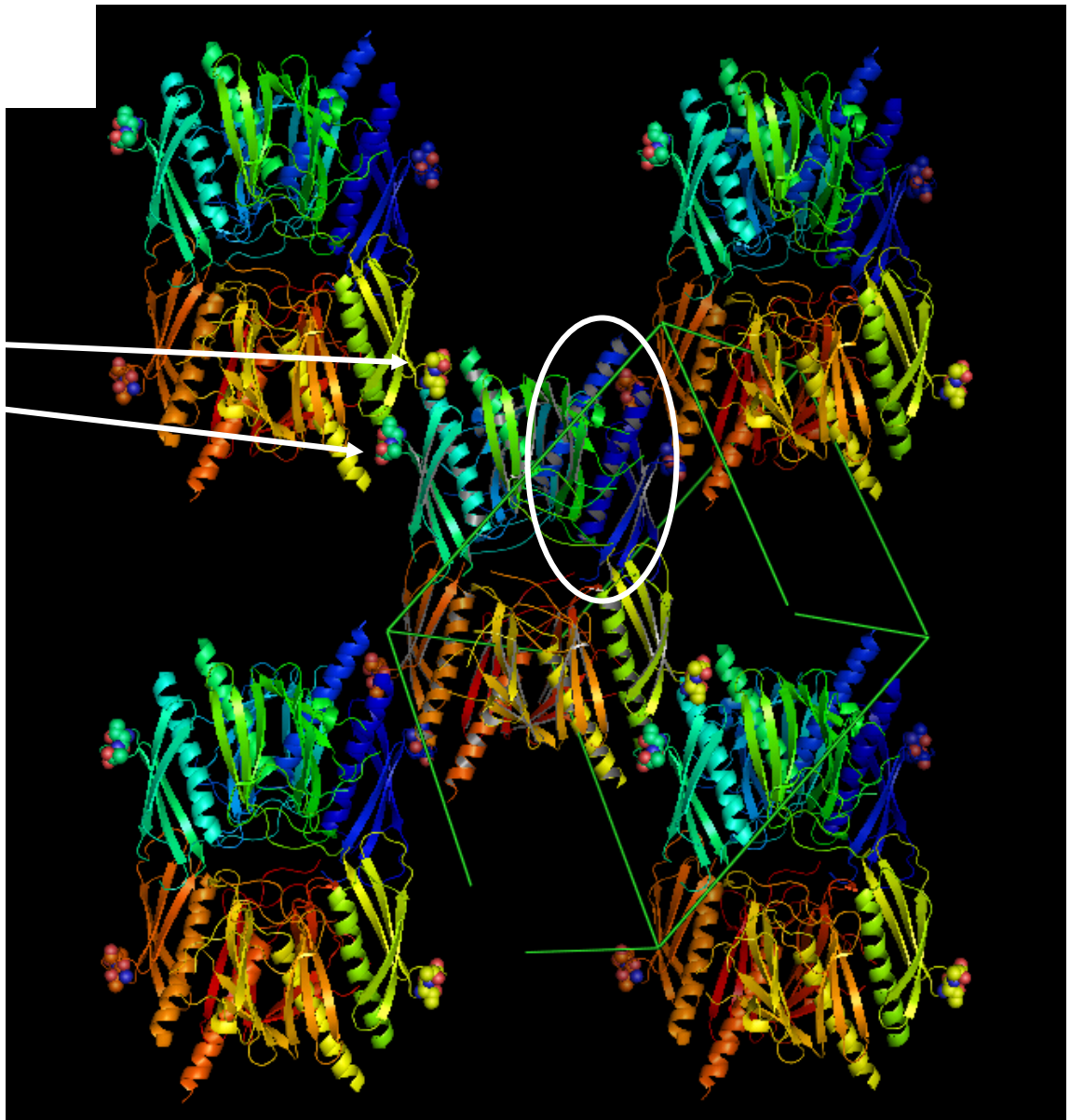
Available online at www.sciencedirect.com

ScienceDirect

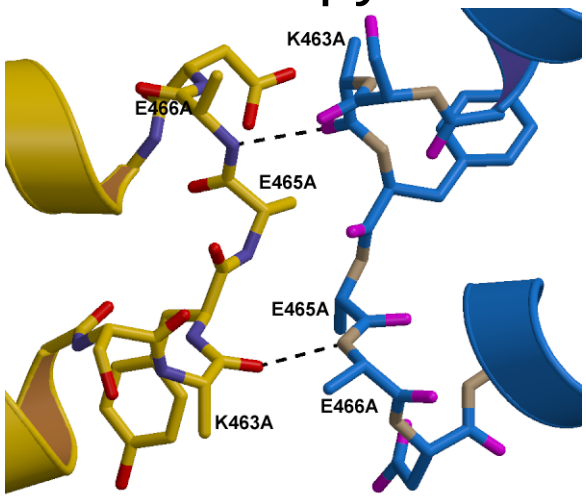
ELSEVIER

The Crystal Structure of a Binary Complex of two Pseudopilins: EpsI and EpsJ from the Type 2 Secretion System of *Vibrio vulnificus*

Marissa E. Yanez[1,2], Konstantin V. Korotkov[1], Jan Abendroth[1] and Wim G. J. Hol[1,2*]

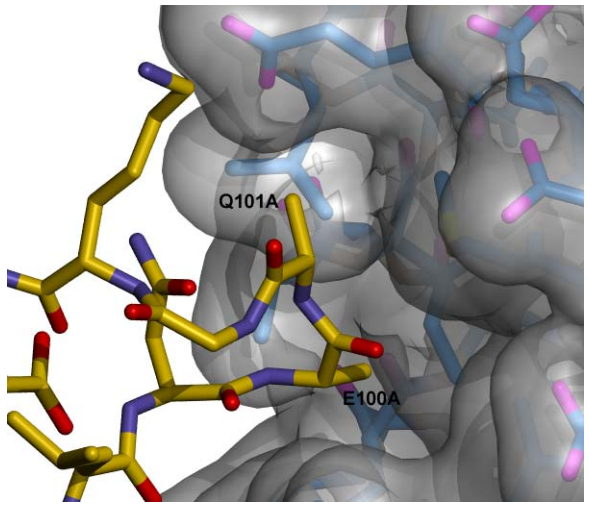


The recurrence of crystal contacts involving mutated sites validates the hypothesis that crystallization is facilitated by surface entropy reduction.



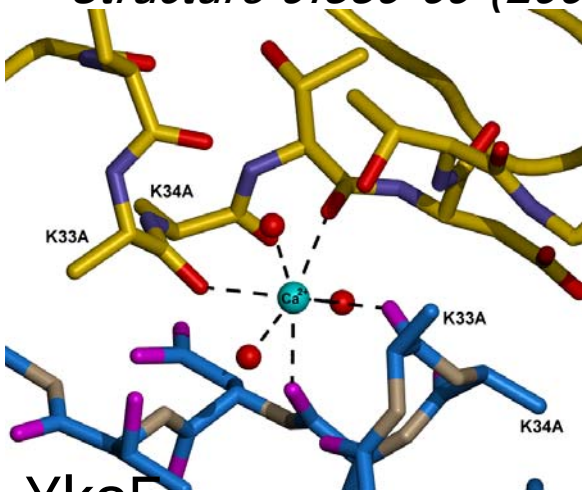
RGSL domain of PDZ-RhoGEF

Structure 9:559-69 (2001)



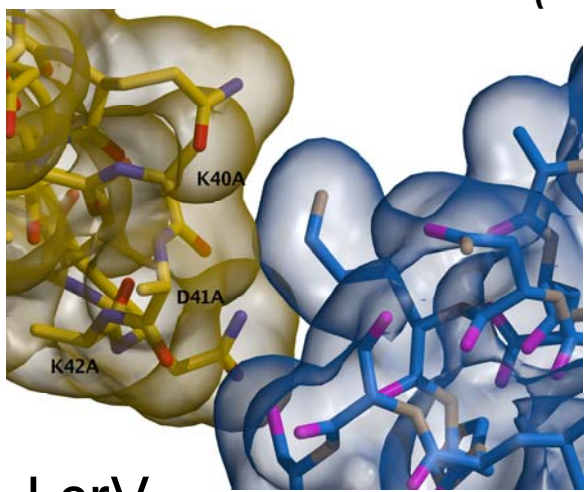
Hsp33

Structure 12:1901-7 (2004)



YkoF

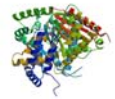
JMB 343:395-406 (2004)

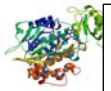


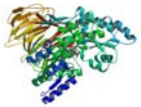
LcrV

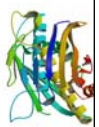
Structure 12:357-8 (2004)

New structures


2fy3, 2fy5, 2fy2, 2fy4

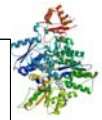

2bra

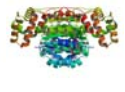


3blc 3dm0



3fms 2bjo

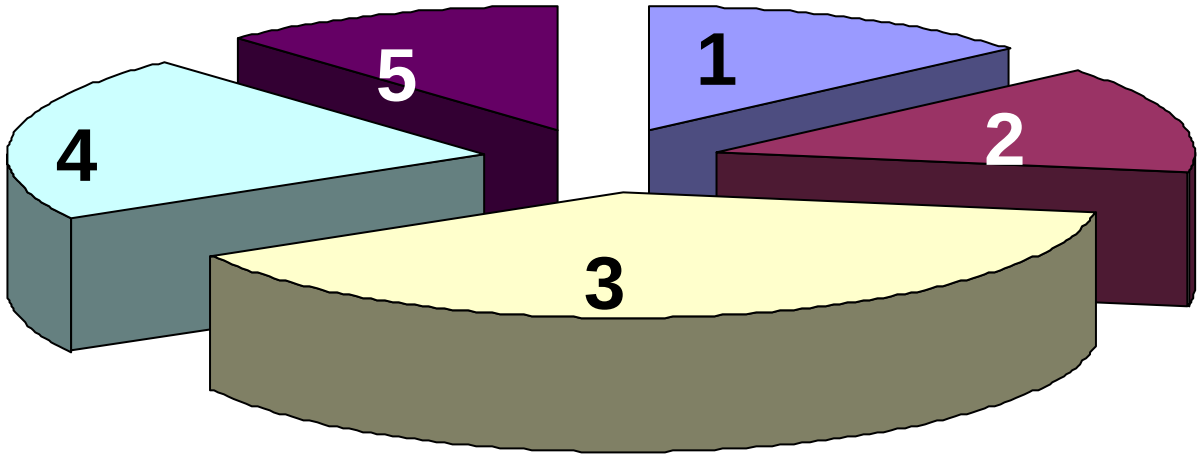


2aj7* 2obn*


3dwg


2jl1,2vrb, 2vrc

How difficult are these proteins for crystallization, as predicted by XtalPred ?



New crystal forms wit


1p4o


1zhp, 1zhr, 1zhm, 2fda,

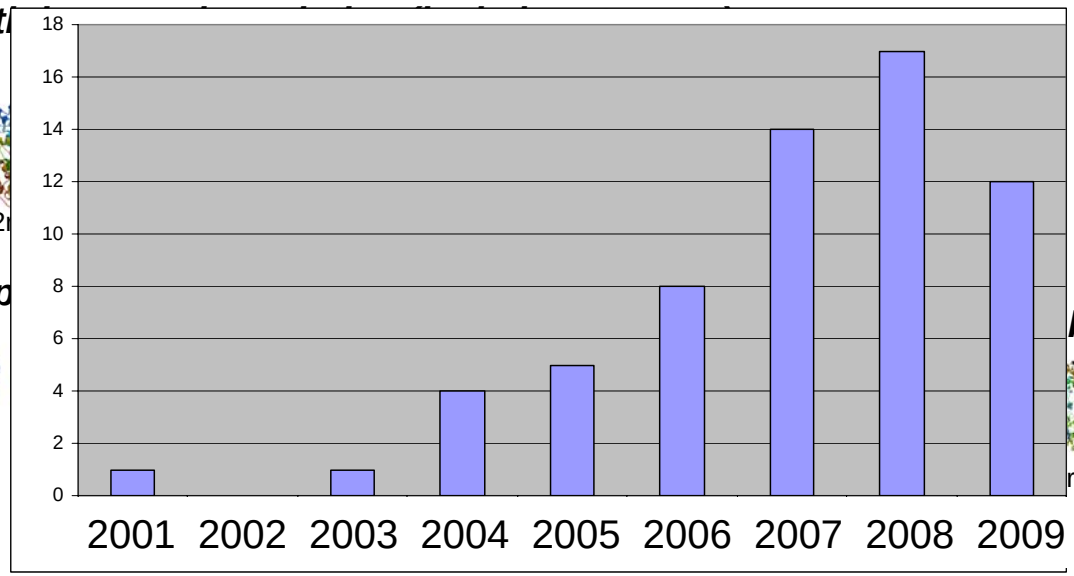
2r

Protein-protein comp

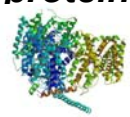

1p3q


3di3, 3di2


2ret



brane proteins


2wit**

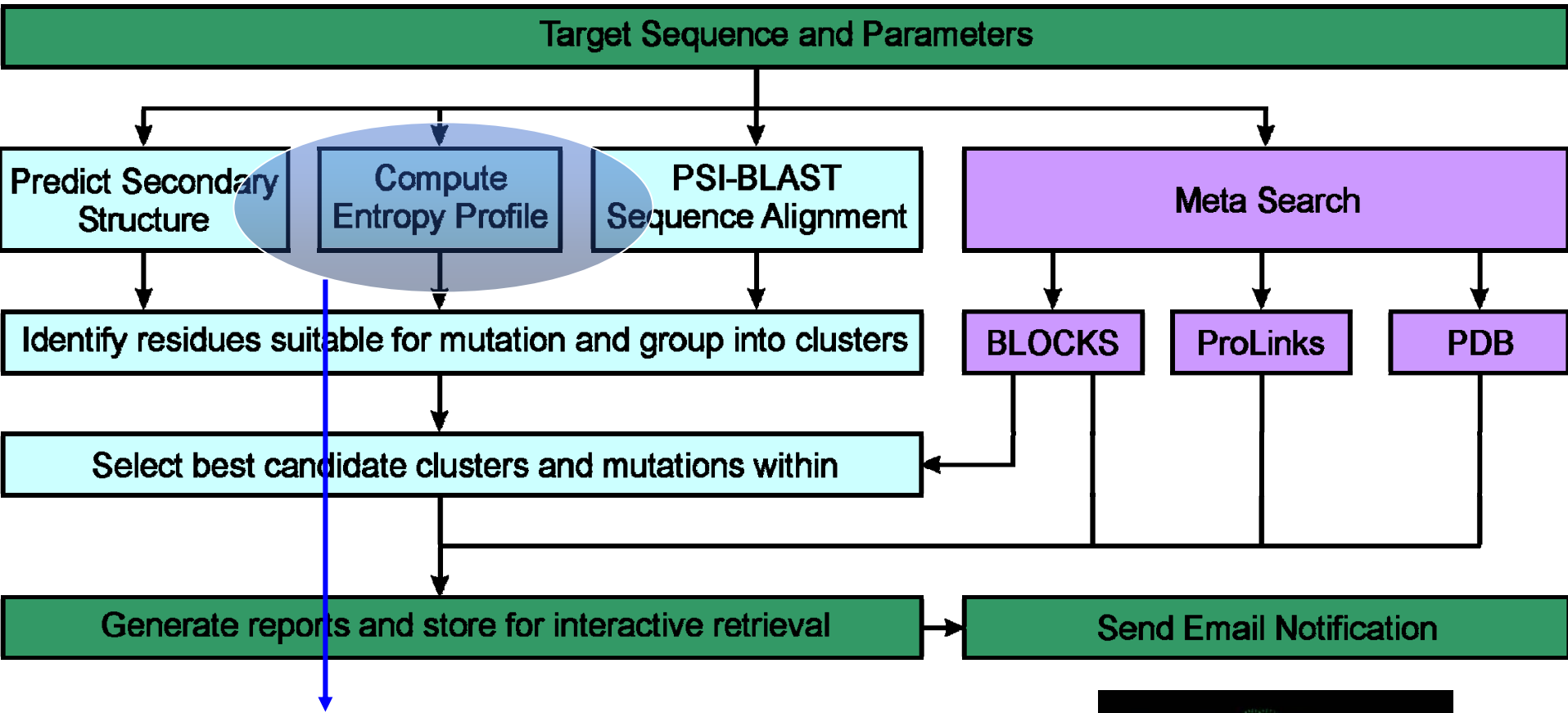
SER Server

http://nihserver.mbi.ucla.edu/SER/

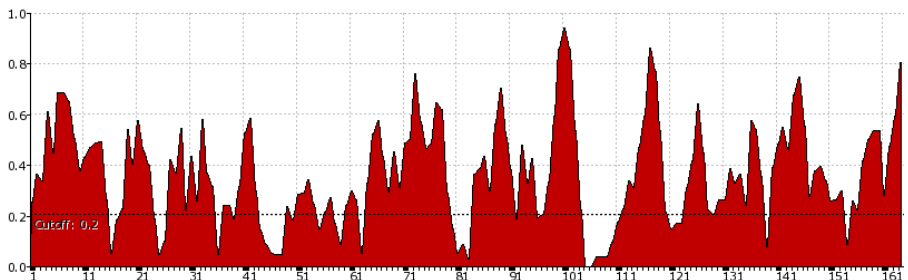
Toward rational protein crystallization: A Web server for the design of crystallizable protein variants

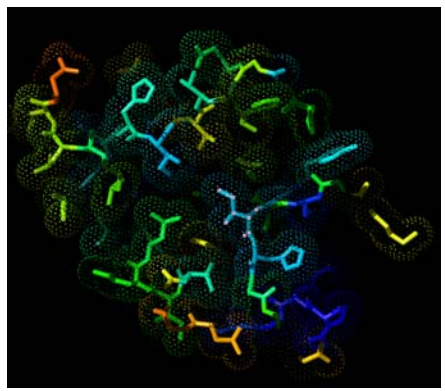
Lukasz Goldschmidt, David R. Cooper, Zygmunt S. Derewenda and David Eisenberg

Protein Sci. 2007 16: 1569-1576



A normalized, smoothed 'free' conformational entropy profile is then computed in a sliding window of three residues





Job & Sequence | **Parameters** | Meta Search | Results

Secondary Structure Prediction - Coil Regions: [?]
- Coil Confidence Cutoff: 0.2 [?]
- Minimum Patch Length: 3 [?]
- Sec. Str. Scoring Weight: 1 [?]

High Entropy Regions: [?]
- Averaging Window: 3 [?]
- Lower Cutoff Limit: 0.2 [?]
- Entropy Scoring Weight: 1.5 [?]

Blast Search: [?]
- Sequence Identity Cutoff: 20 [?]
- Expectation Value Cutoff: 1e-3 [?]
- Conservation Weight: 1 [?]
- Popularity Weight: 0.5 [?]

Cluster & Residue Definitions:
- Min. Mutations per Cluster: 1 [?]
- Max. Mutations per Cluster: 3 [?]
- Max. Gap within Cluster: 2 [?]
- High Entropy Residues: KEQD [?]
- Mutable Amino Acids: KEQ [?]
- Target Amino Acids: ATY [?]

Cluster Scoring Weights:
- Combined residue score: 1 [?]
- Residue score deviation: 1 [?]
- Length of low-entropy patch: 1 [?]
- Gaps in low-entropy patch: 1 [?]
- Number of Mutations: 1 [?]
- Change in Entropy: 1 [?]

Sort Clusters by: [?] (•) Cluster Score ( ) Residue No

Cluster Limit (when sorted by score): [?]
- ( ) No limit show all
- (•) Generate Maximum 10 clusters
- ( ) Cut off at 66 % of max score

Graphical Representation: [?]
- [x] Score Summary
- [x] Blast Conservation
- [x] Entropy average
- [x] Secondary Structure (all)
- [x] Secodary Structure (coil only)

Graph Size: [?]
- ( ) Fixed size: 1000 x 250 pixels
- (•) Relative size: 5 pixels per residue

Update Job | Reload Default Parameters

Proposed Mutations:

Cluster #1:
- K 117 => ATY
- Q 118 => ATY

Residues 117 - 118: KQHY [?]
Score: 9.93 [?]

Cluster #2:
- Q 5 => ATY
- K 7 => ATY
- Q 8 => ATY

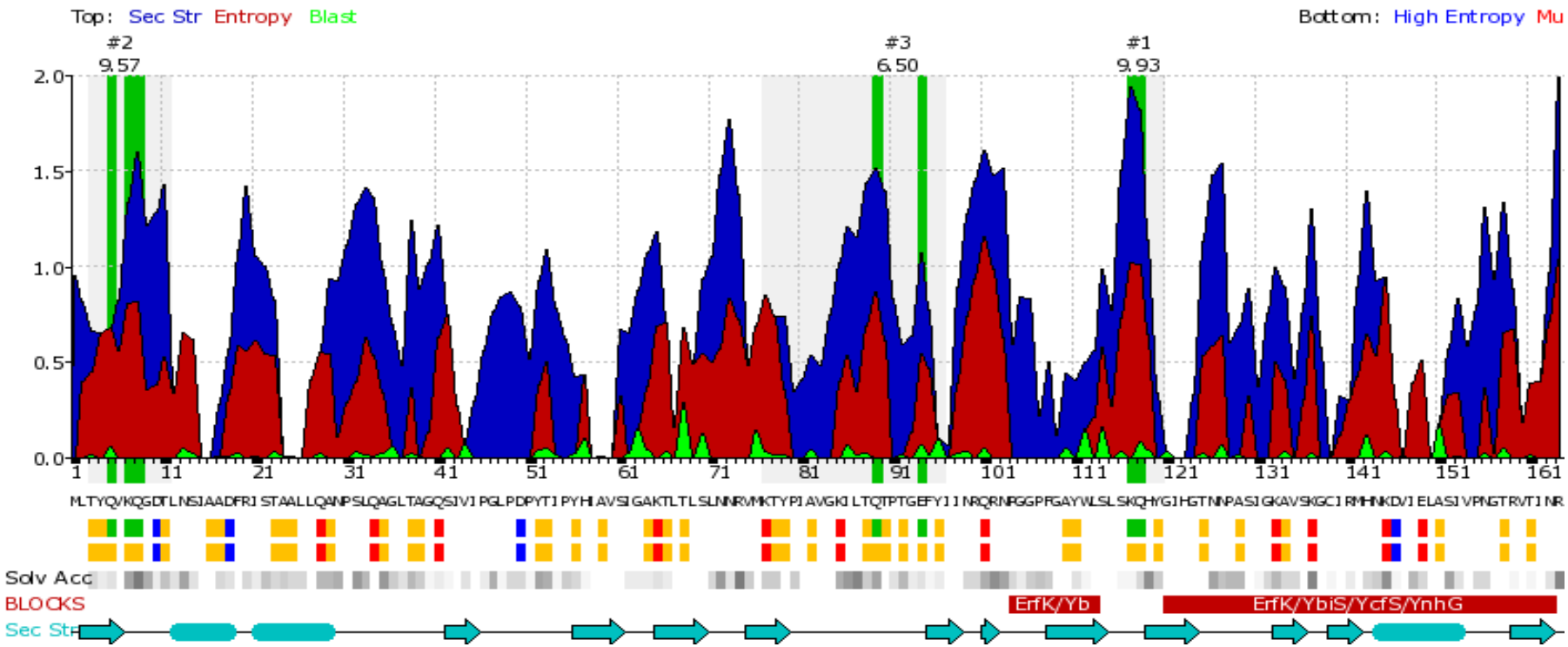
Residues 3 - 8: TYQVKQGDT [?]
Score: 9.57 [?]

Cluster #3:
- Q 89 => ATY
- E 94 => ATY

Residues 88 - 94: KTYPIAVGKILTQTPTGEFY [?]
Score: 6.5 [?]

Cluster #4:
- Q 101 => ATY

Residues 101 - 101: Q [?]
Score: 4.54 [?]



A note of caution:

There is no panacea for all crystallization problems: in many cases various techniques must be synergistically applied until a crystallizable version of the protein is generated, as in the example shown here.

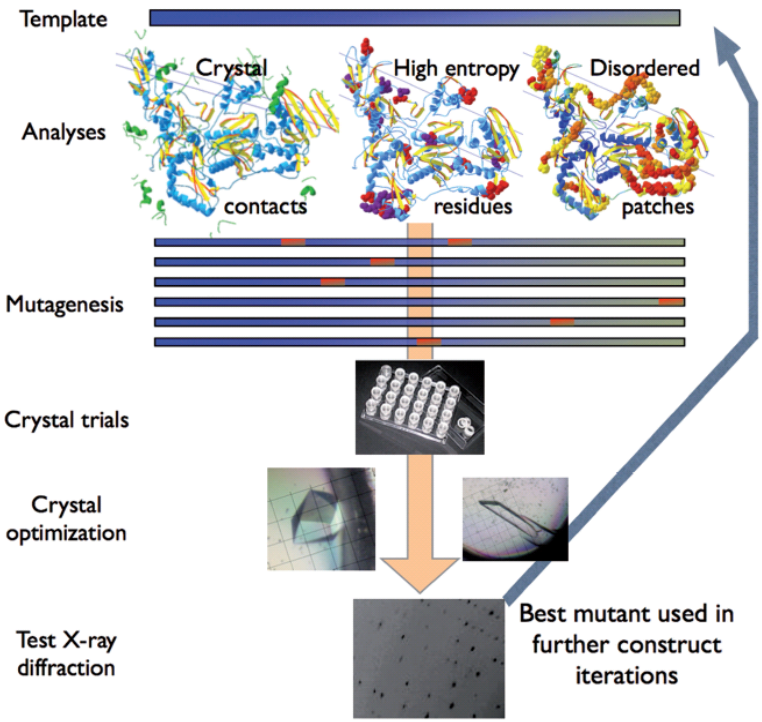
Published online 1 August 2008 *Nucleic Acids Research, 2008, Vol. 36, No. 15* **5083–5092**
doi:10.1093/nar/gkn464

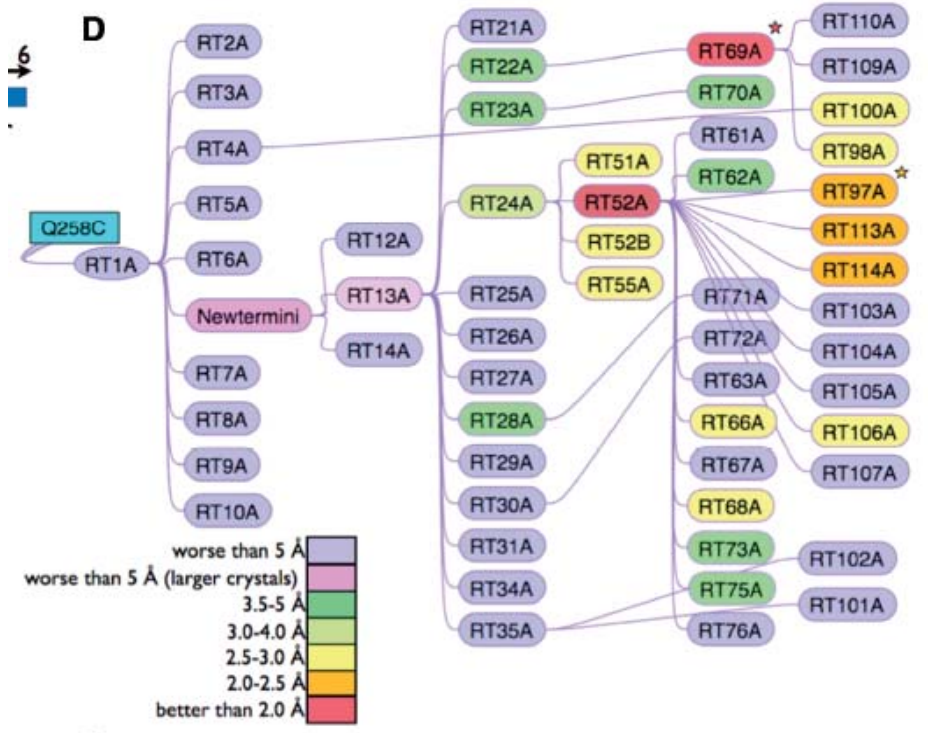
Crystal engineering of HIV-1 reverse transcriptase for structure-based drug design

Joseph D. Bauman[1,2], Kalyan Das[1,2], William C. Ho[1,2], Mukta Baweja[1], Daniel M. Himmel[1,2], Arthur D. Clark Jr[1,2], Deena A. Oren[1,2], Paul L. Boyer[3], Stephen H. Hughes[3], Aaron J. Shatkin[1] and Eddy Arnold[1,2,*]

[1]Center for Advanced Biotechnology and Medicine, [2]Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, NJ and [3]NCI-Frederick Cancer Research and Development Center, Frederick, MD, USA

Received May 1, 2008; Revised July 2, 2008; Accepted July 3, 2008





Acknowledgements:

Past and Present Members of the group:

David Cooper
Urszula Derewenda
Darkhan Utepbergenov
Ankoor Roy
Won-Chan Choi
Meiying Zheng
Yancho Devedjiev
Peter J. Sheffield
Sarah M. Garrard
Kenton Longenecker

Students

Marcin Cieslik*
Agnieszka Mateja*
Jan Czepas*
Aleksandra Gabrys*
Izabela Janda*
Jakub Bielnicki*
Yogesh Surendranath*
Malgorzata Pinkowska*
Katarzyna Grelewska*
Pawel Boczek*
Michal Zawadzki*
Dominika Kamrowska
Justyna Frydrysiak
Grzegorz Czelakowski

Funding: NIGMS/PSI

Collaborators (ISFI):

Tom Terwilliger, LANL
Geoffrey Waldo, LANL
David Eisenberg, UCLA
Luki Goldschmidt
Tony Kossiakoff, University of Chicago
Brent Segelke, LLNL
Li-Wei Hung, LLNL

Collaborators:

Zbigniew Dauter, NCI, APS
Andrzej Joachimiak, ANL, APS
Ian A. Wilson, JCSG
Scott Lesley, JCSG