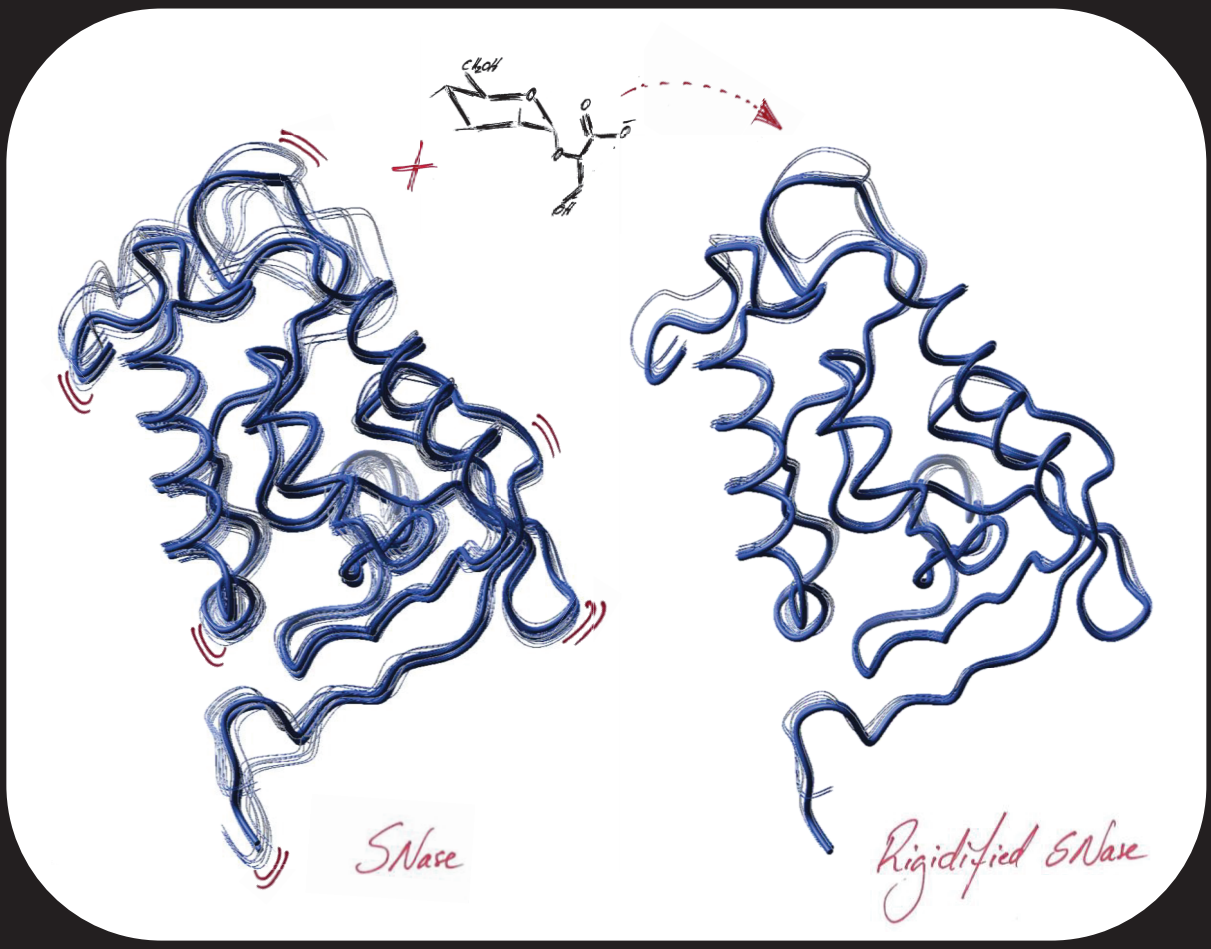
Insights into the molecular mechanisms of protein stabilization by osmolytes of hyperthermophiles

Tiago V. D. Moreira Pais



Dissertation presented to obtain the Ph.D. degree in Biochemistry

Instituto de Tecnologia Química e Biológica | Universidade Nova de Lisboa

Oeiras,
December, 2011

INSTITUTO
DE TECNOLOGIA
QUÍMICA E BIOLÓGICA
/UNL

Knowledge Creation

Insights into the molecular mechanisms of protein stabilization by osmolytes of hyperthermophiles

Tiago Vasconcelos Duarte Moreira Pais

Supervisor: Professor Helena Santos
Co-supervisor: Professor David L. Turner

Dissertation presented to obtain the Ph.D. degree in Biochemistry

Instituto de Tecnologia Química e Biológica | Universidade Nova de Lisboa

Oeiras, December, 2011

FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA EDUCAÇÃO E CIÊNCIA

INSTITUTO DE TECNOLOGIA QUÍMICA E BIOLÓGICA /UNL
Knowledge Creation

Apoio financeiro da Fundação para a Ciência e a Tecnologia (FCT) e do FSE no âmbito do Quadro Comunitário de apoio, Bolsa de Doutoramento SFRH//BD/42210/2007.

In loving memory of my grandparents

From left to right: Eurico Cabrita, Carlos Romão, David Turner, Tiago Pais, Helena Santos, António Maçanita, Rui Brito and Miquel Pons. ITQB, Oeiras 13th December 2011.

ACKNOWLEDGMENTS

As I get to this point I have no doubt that I couldn't have done it without the support of a number of people whom, in different ways and at different times, gave the best of them. However, in the day-to-day life we often don't take the opportunity to express this gratitude. That is why I would like to properly acknowledge all those that made a difference during this long journey.

By far, my supervisor, Professor Helena Santos, had the largest impact in my shaping as a researcher and as a professional, and this will surely be reflected on the rest of my career. With her, discussions are a constant challenge to go further, to excel ourselves, to question the fundamentals, to think differently, to have new ideas and to grow stronger. Most of all, these discussions are an opportunity to identify weaknesses and overcome them, which very often comes at the expense of hard but important times. I also show my appreciation for her never-ending effort in teaching me to communicate ideas, work results and objectives in a way that people other than me are able to understand. Most importantly, I must thank for being allowed to always state my mind. I remember one of the first meetings where I was taught that there are no silly questions, only silly answers. And if at any time I have become short of my own goals I can never ascribe it to lack of working conditions. Professor Helena Santos has always provided the group with the best working conditions. For all of this I am grateful as I am convinced to have become a better scientist, more resourceful, demanding, accurate, aware and more confident in overcoming the challenges to come. Of course, the rigour, the scientific knowledge, the enthusiasm and the resilience to the challenging scientific Portuguese reality that she has shown over the years can only inspire any young researcher to go ever further.

To Professor David Turner, my co-supervisor, with whom I have learned so much about biophysical subjects, I need to express my appreciation for his commitment in providing always the sharpest analysis about my work, for pushing me to go deeper

into the subjects, and for teaching me to be critic about pre-established ideas and concepts.

To Dr. Pedro Lamosa for introducing me to the NMR world, for showing trust in me since very early, and for his tremendous capacity of quickly analysing problems and suggesting ideas to overcome them. I am also thankful to him for having introduced to me the world of extremophyles with such an enthusiasm that immediately captured my curiosity.

To Dr. Nuno Borges for always making me feel welcome even when my moods were not so pleasant. I thank him also for being constantly available and willing to help at all times and unconditionally. I also appreciate very much his effort of integrating me in the group, with frequent requests to help him about his own doubts – although most of the times I was fairly sure he knew the answers already. I must not forget to thank him also for the enourms dedication and effort that he has put for the last years in keeping the lab a proper place for conducting quality scientific work.

To Dr. Luis Gafeira for his contagious enthusiasm in discussing science, for his bright analysis of my work, for always showing interest and making the effort to understand even the most complicated data.

Pedro, Nuno and Luis are like the three muskteers of the group, always willing to come to the rescue; I am very lucky for having them as friends.

To Dr. Marta Rodrigues for her unconditional friendship, for her outstanding ability to bring warmth to the lab and for never giving up on friends. Her commitment with true science, with apparently unimportant details and with never taking shortcuts is a stimulus for all who wish to make big findings.

To Dr. Carla Jorge for being always an honest colleague, for keeping the scientific level of research at all times high, and for being a concerned friend.

To Dr. Tony Collins for quickly becoming an excellent colleague and a good friend. I thank him also for the renewed enthusiasm that he brought to the lab, with a clear will of giving the best of him to the group and to his research work.

To Dr. Manolis Matzapetakis for showing me every trick he knows about NMR without any conditions. I thank him also for priceless critic analysis of my work, for his concerns about my future and for good advices – aim high!

To Dr. Tiago Faria, a unique colleague who always had the capacity to demonstrate a true belief in my work. For that I thank him deeply because, I have to admit, many times I failed to believe in my own work. Even now, despite having moved to a different lab at Coimbra, he is still able to inspire trust in my competences.

To Dr. Paula Fareleira for always showing support, for being an example of resourcefulness and for rapidly treating me as a true colleague despite her senior condition. I also thank her effort for promoting a friendly environment, especially outside the lab.

To Ana Esteves, the most patient "next door" neighbour a colleague can have. To Ana Mingote, Patricia Almeida and Ana Lúcia for their dedication to the lab and to others group members. To Dr. Teresa Catarino for the bright youthful spirit she can bring to the hardest discussions. To Dr. Luis Fonseca, for helpful insights into statistical analysis. To Dusica Rados, Cristiana Faria, Marta Conchinha, and Pedro Quintas for keeping a friendly and tolerant working atmosphere.
To Laura Paixão, Sónia Neto, Teresa Maio for their joyful spirits which greatly contributed to overcome some less positive days. To Dr. Rute Neves for always sharing her wonderful holydays at remote places. To former members of the group namely Dr. Melinda Noronha (a very bright scientist), Dr. Maria Manuel Sampaio who gave me the first real contact with molecular biology, Dr. Margarida Santos, Filipa

Cardoso, Dr. Rute Castro, Dr. Claudia Sanchez, Dr. Paula Gaspar, Sandra Carvalho, and João Cavalheiro.
To Anabela Bernardo, a very dedicated friend, capable of going a step further to help everyone and in particular her friends. Because of her, life in the lab was several orders of magnitude smoother. To the maintenance crew who always tried their best to deal with problems that insisted in showing up after-hours (very late hours sometimes).

To Fundação para a Ciência e a Tecnologia for the financial support provided by the Ph.D. grant, and to Instituto de Tecnologia Química e Biólogica, for providing conditions to pursue scientific excellence.

Aos meus outros amigos, fora do circulo cientifico, mas sem os quais tenho a certeza que não teria mantido a capacidade para enfrentar os muitos desafios que estes estudos me lançaram. Em particular a pessoas como o Rogério Morais, o Pedro Andrade e Sousa, o Nuno Gomes, o Diogo Serrano, o David Marçal, o Pedro Marques nas quais confiei literalmente a minha vida e que nunca me falharam; aos meus amigos da Serra especialmente o Zé Maria, o José Amoreira, o Rómulo Carvalho, o Filipe Saraiva, o Rui Ribeiro e o Siggi que me alimentam o sonho de manter a minha Serra um lugar de sonho; e ao meu amigo de longa data, Nuno Carvalho que apesar de longe sempre se soube manter próximo.

Por ultimo, mas naturalmente não menos importante, agradeço à minha familia que é o meu porto-seguro em particular à minha mãe, ao meu pai, por serem pessoas extraordinárias e com uma força invejável, à Patita pelo apoio incondicional, à minha avó pelo seu espirito jovem sempre interessado nas minhas investigações e à Rita por todo o carinho e pelo apoio ao longo destes anos.

ABSTRACT

During the last decades, the discovery of microorganisms thriving in extremely inhospitable habitats totally changed our view on the limits for Life. Extremophiles have been isolated from scalding water springs and from permanently frozen soil, from hypersaline Dead Sea samples as well as from highly acidic or alkaline sites. However, there is no evidence for the occurrence of life in the absence of water. Cell membranes are permeable to water, hence it is a prerequisite for survival that cells can cope with changes in water activity. Most cells are prepared to counteract an eventual increase in the external osmotic pressure by accumulating small molecules designated osmolytes in their cytosol. Organisms adapted to hypersaline environments accumulate molar-amounts of osmolytes, such as glycerol or KCl. Despite the huge concentrations, these compounds do not interfere with the normal functions of macromolecules in the cell, thereby earning the designation of "compatible solutes". Interestingly, the role of compatible solutes goes well beyond that of balancing the osmotic gradient across the cell membrane and actually some of these compounds are involved in the response to other types of stress, such as heat, oxidative and acid stress. Conversely, typical osmolytes, such as trehalose or ectoine, have been shown to confer increased stability to proteins under a range of environmental challenges. Our group directed a great research effort to study compatible solutes from organisms adapted to hot environments (hyper/thermophiles), and to investigate their contribution to the stabilization of proteins against heat denaturation. Organic osmolytes and other compatible solutes fall into a few categories of chemical compounds, namely amino acids and derivatives, polyols, sugars and derivatives. Interestingly, marine organisms that thrive at high temperatures accumulate specific solutes that usually bear a negative charge at

physiological pH and are rarely found in mesophilic organisms. This is the case of mannosylglycerate and diglycerol phosphate, two compatible solutes highly restricted to (hyper)thermophiles and which have been selected as objects of study in this thesis. While diglycerol phosphate is restricted to the domain Archaea, mannosylglycerate can also be found in members of the Eukarya and Bacteria.

Understanding the molecular mechanisms that govern protein stabilization by osmolytes and other compatible solutes has become a subject of great current interest, not only due to the intrinsic biotechnological significance of increasing the stability of enzymes under operational conditions, but also because of the clinical relevance of finding effective suppressors of protein misfolding and aggregation. In fact, the increasing prevalence of neurodegenerative diseases and other pathologies associated with protein misfolding and amyloid formation has attracted further attention to the classical studies of osmolyte-induced protein protection. Naturally, the long term goal is the rational design of optimal stabilizers for specific target proteins.

The work described in this thesis aims at understanding the molecular interactions that contribute to the stabilization of the protein native state by compatible solutes typical of hyper/thermophiles. In particular we set out to identify structural determinants of protein stabilization and to gain insight into the importance of changes in protein dynamics to the mechanism of solute induced stabilization. Rubredoxin from *Desulfovibrio gigas* and staphylococcal nuclease were the two target model proteins used in this study. NMR was the methodology chosen to assess changes in protein motions in a wide range of times-cales and eventual binding to specific features of the protein structure.

Evidence for the involvement of structural determinants comes from previous studies in our group showing that rubredoxins from closely related

organisms and with high structural similarity are stabilized to very different extents by diglycerol phosphate. The size of a hairpin structure that partly protects the hydrophobic core from solvent access is the main structural difference between the two homologues examined. Therefore, the length of such loop was manipulated in the rubredoxin of *Desulfovibrio gigas* to assess the importance of this structural feature to protein stabilization by compatible solutes. The three-dimensional structure of the mutant with the shortest loop, Δ17|29, was determined by NMR, demonstrating that, despite the drastic deletion (13 residues), the main structural features of the native protein were preserved. The kinetic stability of the mutants was assessed at 90°C by monitoring the absorbance decay at 494 nm that results from the metal centre disruption and consequent iron release. It was found that the extent of kinetic stabilization conferred by diglycerol phosphate and mannosylglycerate was strongly dependent on the specific mutant examined. In the case of diglycerol phosphate, the differences were dramatic and the half-life for iron release of mutant Δ23|29 was decreased in contrast to the threefold increase observed for the *D. gigas* native rubredoxin. Chemical shift analysis indicated that both solutes establish preferential interactions with residues located at the loop region of the rubredoxin mutant Δ17|29, and that both solutes induce a generalised increase in the strength of the hydrogen bond network which may imply a more compact protein structure. Thus, this work provided evidence for the loop as being a structural determinant of rubredoxin stabilization and for weak, specific interactions of the solutes with the protein surface.

The other goal in this work was to clarify the relationship between protein motions and protein stabilization. Prior to this thesis, measurements of hydrogen/deuterium exchange rates of amide protons suggested that the global unfolding events that enable the exchange of the most protected -NH groups, were notably restricted by stabilizing solutes. However, information

about the effect of stabilizing solutes on motions at faster timescales was very limited and the structural detail needed to determine specific effects along the amino acid sequence was lacking. Moreover, all studies referred to canonical solutes of mesophiles, *i.e.*, non-charged molecules. Therefore, we set out to perform a detailed study of the effect of mannosylglycerate on the various timescale motions of a hyperstable staphylococcal nuclease variant by using multiple NMR spectroscopy methods.

Sub-nanosecond motions of the backbone and side-chains were studied, respectively, by ^{15}NH and $^{13}CH_3$ relaxation measurements. Model-free analysis of the relaxation data showed that backbone and side-chain motions were differently affected by the presence of mannosylglycerate (up to 0.35 M concentration). At high frequencies, internal motions of staphylococcal nuclease were progressively restricted with increasing concentrations of mannosylglycerate or reduced temperature, while the opposite effect was observed with 0.25 M urea (a destabilizing solute). Importantly, the order parameters showed a strong correlation with the changes in the T_m values induced by different solutes or concentrations, as determined by differential scanning calorimetry. In contrast, side-chain motions were little affected by any of the solutes tested, and showed no correlation with changes in the thermal stability of staphylococcal nuclease. Our data show that mannosylglycerate caused a generalized reduction of backbone motions in the sub-nanosecond timescale and make apparent a correlation between protein stabilization and backbone rigidification.

The widest and slower movements of the protein (second to minute time scale) were significantly constrained in the presence of mannosylglycerate, as indicated by the strong decrease of the hydrogen/deuterium exchange rates, but on the millisecond timescale the effect was much smaller and comparable to that of glycerol, a solute with no impact on the T_m of the protein.

Furthermore, the analysis of the urea concentration dependence of –NH exchange rates allowed for the differentiation of amide protons involved in local and global protein motions. Thus, it was possible to conclude that mannosylglycerate restricted the slower local motions in addition to the global motions of the protein. Importantly, the local motions of the β-sheet residues were specifically restricted by the presence of mannosylglycerate. On the other hand, the effect of mannosylglycerate on the global motions was not structure specific; this solute appeared to simply cancel out the deleterious effect of urea in a proportion of one part of mannosylglycerate to 3.5 parts of denaturant. To our knowledge this is the first time that a change in the motions of a specific structural element is ascribed to the effect of a stabilizing solute.

In summary, the work presented in this thesis shows that structural features can be a source of specific interactions between solutes and proteins which may relate with the differential extent of stabilization observed for specific solute-protein couples. A correlation between protein stabilization and protein backbone rigidification of motions in the sub-nanosecond timescale is clearly established for the model system analyzed. Finally, in the second to minute timescale, there is strong evidence for solute-induced restriction of the local as well as the global motions of the protein. Further studies will show whether these interesting findings can be generalized to other proteins and solutes.

RESUMO

A descoberta de microrganismos que proliferam em habitats extremamente inóspitos mudou totalmente a visão da comunidade científica sobre as condições limites para ocorrência de Vida. Têm sido isolados organismos extremófilos a partir de nascentes de água fervente e também de amostras de solos permanentemente gelados, a partir de amostras hipersalinas do Mar Morto, bem como de ambientes extremamente ácidos ou alcalinos. No entanto, não há qualquer evidência para ocorrência de vida na ausência de água. As membranas celulares são permeáveis a água, por isso a sobrevivência da célula implica necessariamente a sua capacidade para reagir adequadamente a alterações na actividade da água no meio exterior. A maioria das células está preparada para contrabalançar um eventual aumento na pressão osmótica externa com a acumulação no citoplasma de moléculas de baixa massa molecular que se designam osmólitos. Organismos adaptados a ambientes hipersalinos acumulam concentrações elevadas de osmólitos, tais como glicerol ou KCl. Apesar das concentrações poderem atingir valores acima de 1 M, estes compostos não afectam negativamente as funções das macromoléculas na célula, sendo por isso designados "solutos compatíveis". É interessante notar que o papel dos solutos compatíveis ultrapassa grandemente a mera função de equilibrar o gradiente osmótico através da membrana celular; de facto, alguns desses compostos estão envolvidos na resposta a diferentes tipos de agressão, tais como agentes oxidantes, calor ou pH baixo. Além disso, foi demonstrado que osmólitos típicos, tais como trealose ou ectoina, são capazes de proteger a estrutura de proteínas e de outras macromoléculas contra uma variedade de agressões ambientais.

A nossa equipa de trabalho tem-se empenhado no estudo dos solutos compatíveis acumulados por organismos adaptados a ambientes quentes,

(hiper)termófilos, e na elucidação dos mecanismos moleculares subjacentes aos efeitos de estabilização de proteínas contra a desnaturação térmica. Osmólitos orgânicos, e outros solutos compatíveis, enquadram-se num número restrito de categorias de compostos químicos, nomeadamente aminoácidos e derivados, polióis, açúcares e derivados. Curiosamente, os microrganismos de origem marinha e que proliferam a altas temperaturas acumulam solutos específicos que geralmente têm carga negativa ao pH fisiológico e raramente ocorrem em organismos mesofílicos. É o caso de manosilglicerato e de fosfato de diglicerol, dois solutos compatíveis tipicamente encontrados em (hiper)termófilos e que foram escolhidos como objecto de estudo nesta tese. Enquanto o fosfato de diglicerol se restringe ao domínio *Archaea*, o manosilglicerato também pode ser encontrado em membros dos domínios *Eukaria* e *Bacteria*.

Compreender os mecanismos moleculares que controlam a estabilização de proteínas por osmólitos e outros solutos compatíveis tornou-se um assunto de grande actualidade, não só devido à importância biotecnológica intrínseca de aumentar a estabilidade de enzimas em condições operacionais, mas também por causa da relevância clínica de encontrar supressores eficazes de erros no enrolamento de proteínas e consequente agregação. Na verdade, a prevalência crescente de doenças neurodegenerativas e de outras patologias associadas ao enrolamento incorrecto de proteínas tem atraído atenção redobrada aos estudos clássicos de protecção de proteínas por osmólitos. Naturalmente, o objectivo a longo prazo é o desenho racional de estabilizadores ideais para proteínas-alvo específicas.

O trabalho descrito nesta tese visa compreender as interacções moleculares que contribuem para a estabilização do estado nativo de proteínas por solutos compatíveis típicos de organismos (hiper)termófilos. Em especial, propusémo-nos identificar os determinantes estruturais da

estabilização de proteínas e avaliar a importância de alterações na dinâmica das proteínas para o mecanismo de estabilização. A rubredoxina de *Desulfovibrio gigas* e a nuclease de *Staphylococcus aureus* foram as proteínas-modelo utilizadas neste estudo. A Ressonância Magnética Nuclear foi a metodologia escolhida para avaliar as alterações nos movimentos internos da proteína, numa ampla gama de escalas de tempo, e a possibilidade de ligação a elementos estruturais específicos da proteína.

Evidência para o envolvimento de determinantes estruturais surgiu de estudos anteriores do nosso grupo mostrando que o fosfato de diglicerol conferia graus de estabilização muito diferentes a duas rubredoxinas com elevada semelhança estrutural e provenientes de organismos muito próximos em termos de filogenia. A diferença estrutural mais marcante entre as duas proteínas homólogas examinadas é o tamanho de um elemento estrutural tipo gancho de cabelo que protege parcialmente o núcleo hidrofóbico do acesso ao solvente. Assim sendo, o tamanho desta estrutura foi manipulado na rubredoxin de *Desulfovibrio gigas* para se avaliar a relevância desta característica estrutural na estabilização da proteína por solutos compatíveis. A estrutura tridimensional do mutante com o menor tamanho do elemento tipo gancho, Δ17|29, foi determinada por RMN, demonstrando que, apesar da eliminação drástica de 13 aminoácidos, o mutante exibia as características estruturais essenciais da proteína nativa. A estabilidade cinética dos mutantes foi avaliada a 90ºC, monitorizando o decaimento da absorvância a 494 nm que resulta da ruptura do centro metálico e consequente libertação de ferro. Verificou-se que o grau de estabilização cinética conferido por fosfato de diglicerol e por manosilglicerato era extremamente dependente do mutante especifico examinado. No caso de fosfato de diglicerol as diferenças foram enormes, tendo-se verificado uma redução do tempo de meia-vida para a libertação de ferro no mutante Δ23|29 em contraste com o aumento de três

vezes observado com a rubredoxina de *D. gigas*. A análise dos desvios químicos dos sinais da proteína, na presença e ausência de solutos, indicou que ambos os compostos estabelecem interacções preferenciais com resíduos localizados na região da estrutura em gancho do mutante Δ17|29 e causam um aumento generalizado na coesão da rede de ligações por pontes de hidrogénio, o que sugere compactação da estrutura da proteína. Em suma, este trabalho identificou um elemento tipo gancho de cabelo como sendo um determinante estrutural na estabilização de rubredoxinas e evidenciou interacções específicas fracas dos solutos com a superfície proteica.

Um segundo objectivo neste trabalho foi o de esclarecer a relação entre os movimentos internos das proteínas e a estabilização conferida por solutos de (hiper)thermófilos. Antes deste trabalho de tese, a medição das taxas de permuta hidrogénio/deutério de grupos amida sugeriu que os movimentos globais de desenrolamento, que possibilitam a troca dos protões mais protegidos, estavam consideravelmente restringidos na presença de solutos estabilizadores. No entanto, a informação sobre o efeito de solutos estabilizadores nos movimentos das proteínas em escalas de tempo mais rápidas era muito escassa e faltavam os detalhes estruturais necessários para determinar eventuais efeitos específicos ao longo da cadeia de aminoácidos. Além disso, os estudos disponíveis referiam-se exclusivamente a solutos canónicos de mesófilos, ou seja, moléculas sem carga eléctrica. Por isso, propusémo-nos realizar um estudo detalhado do efeito de manosilglicerato nos movimentos de uma variante superestável da nuclease de *Staphylococcus aureus*, numa gama larga de escalas de tempo, usando diversos métodos de espectroscopia de RMN.

Os movimentos muito rápidos - escala de sub-nanosegundos - do esqueleto proteico e das cadeias laterais foram estudados por medidas de taxas de relaxação de ^{15}NH e $^{13}CH_3$, respectivamente. A análise dos dados de relaxação

pelo método "Model-free" mostrou que os movimentos do esqueleto e das cadeias laterais da proteína foram afectados diferentemente pela presença de manosilglicerato (até 0,35 molar de concentração). Na gama das altas frequências, os movimentos internos da nuclease foram progressivamente restringidos para concentrações crescentes de manosilglicerato ou redução gradual da temperatura, enquanto o efeito oposto foi observado com 0,25 M de ureia (soluto desestabilizador). De notar que os parâmetros de ordem mostraram uma forte correlação com as alterações nos valores da temperatura de fusão da proteína, induzidas por diferentes solutos ou concentrações, como determinado por calorimetria diferencial de varrimento. Pelo contrário, os movimentos das cadeias laterais foram pouco afectados por qualquer dos solutos testados, não existindo evidência de correlação com alterações na estabilidade térmica da nuclease. Os nossos dados mostram que o manosilglicerato induz uma restrição generalizada dos movimentos do esqueleto da proteína na escala dos sub-nanosegundos e evidenciam uma relação directa entre a estabilização da proteína e a rigidificação do respectivo esqueleto proteico.

Os movimentos mais amplos e lentos da proteína (escala de tempo de segundo a minutos) foram significativamente restringidos na presença de manosilglicerato, como evidenciado pela forte diminuição das taxas de permuta hidrogénio/deutério, mas na escala de tempo dos milissegundos o efeito foi muito menor e comparável ao do glicerol, um soluto que não altera o ponto de fusão da proteína. Adicionalmente, a análise da dependência das taxas de permuta dos grupos –NH com a concentração de ureia permitiu distinguir os protões amida que se tornavam acessíveis ao solvente através de movimentos locais ou de movimentos globais. Desta forma, foi possível concluir que o manosilglicerato restringiu não só os movimentos globais da proteína mas também os movimentos locais. É de salientar que os

movimentos locais dos resíduos das folhas beta foram especificamente restringidos pela presença de manosilglicerato. Por outro lado, o efeito do manosilglicerato sobre os movimentos globais da proteína não mostrou especificidade estrutural; este soluto parece simplesmente cancelar o efeito pernicioso da ureia quando na proporção de uma parte de manosilglicerato para 3,5 partes de ureia. Tanto quanto nos é possível saber, esta é a primeira vez que uma alteração nos movimentos de um elemento estrutural específico é atribuída ao efeito de um soluto estabilizador.

Em suma, o trabalho apresentado nesta tese mostra que há elementos estruturais que podem estar na origem de interacções específicas entre solutos e proteínas, as quais por sua vez podem estar relacionados com os diferentes graus de estabilização conferidos consoante o par proteína-soluto em questão. Foi também possível estabelecer uma correlação clara entre a estabilização de proteínas e a restrição dos movimentos do esqueleto proteico na escala de tempo dos sub-nanossegundos. Finalmente, na escala de tempo dos segundos a minutos, existe evidência credível para restrições dos movimentos locais bem como dos movimentos globais da proteína, induzidas pelo soluto. Seria interessante alargar a gama de proteínas objecto de estudo e avaliar a universalidade destas conclusões.

Contents

ABBREVIATIONS

AIC	Akaike's information criteria
AMS	Accessible molecular surface
BMRB	Biological magnetic resonance data bank
CD	Circular dichroism
CLEANEX-PM	Phase-modulated CLEAN chemical exchange
COSY	Homonuclear shift correlation spectroscopy
CPMG	Carr-Purcell-Meiboom-Gill
CSA	Chemical shielding anisotropy
DGP	Diglycerol phosphate
DSC	Differential scanning calorimetry
FlAsH	Fluorescein arsenical hairpin binder
FRET	Förster resonance energy transfer
IDP	Intrinsically disordered protein
HSQC	Heteronuclear single quantum coherence
MG	Mannosylglycerate
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effect
NOESY	Nuclear Overhauser effect spectroscopy
PDB	Protein data bank
RCSB	Research collaboratory for structural bioinformatics
Rd	Rubredoxin
RdDg	Rubredoxin from *Desulfovibrio gigas*
ReAsH	Red arsenical hairpin binder
RMSD	Root mean square deviation
SAXS	Small angle X-ray scattering
SNase	Staphylococcal nuclease

TMAO	Trimethylamine *N*-oxide
TOCSY	Total correlation spectroscopy

Chapter I

General Introduction

CONTENTS

PROTEIN STABILITY

In practical terms, protein stability is the capacity of the polypeptide chain to hold a specific three dimensional average configuration that allows it to perform a given function. This configuration is essentially encoded in the amino acid sequence (Anfinsen 1973) which determines the different types of energetic interactions that constantly drive the protein to its folded functional state. However, not all amino acid residues share the same importance in the definition of the native structure, as demonstrated by proteins with similar native structure but low sequence homology. A multitude of environmental factors challenges the endurance of the folded protein state namely temperature, pH extremes, chaotropic agents, and pressure. In thermodynamic terms, protein stability is described as the energy difference between the native and unfolded states at equilibrium. The unfolding equilibrium constant, *K*, relates to the standard free energy change by the Lewis equation:

$$\Delta G^{\circ} = -RT \cdot lnK \qquad \textbf{[I.1]}$$

where *R* is the gas constant and *T* is the temperature in Kelvin. A free energy change in the range 5-10 kcal/mol applies to most proteins (Jaenicke 1996; Pace et al. 1996). This small number results from two large opposing contributions. On one hand, the enthalpic gain associated with protein folding as a result of the several intramolecular interactions (of the order of hundreds of kcal/mol), and on the other there is a high entropy loss in the protein conformation. A complete energetic balance of the system must also consider the contributions from the solvent molecules. The intricate balance of folding and unfolding forces makes protein stability very difficult to predict even when both the amino acid sequence and the three dimensional structure are known in detail. In fact, comparison of the amino acid content and secondary

structural elements of proteins derived from organisms adapted to different temperatures has yielded, so far, no general rule for greater protein stability (Yip et al. 1995; Szilagyi and Zavodszky 2000; Pfeil et al. 1997; Cambillau and Claverie 2000). Also, recent studies on directed evolution of protein stability showed that increased stability can be achieved in many ways and underlie the difficulty in rationalizing the structural determinants of protein stability (Palackal et al. 2004; Garrett et al. 2004; Englander et al. 2007).

Protein stability can also be analyzed in terms of kinetic stability, which is generally defined as the time course of function loss and is related to the free energy barrier between folded and unfolded states. This property allows proteins to maintain function even when the active state is thermodynamically unstable. Kinetic stability is particularly relevant *in vivo* since aggregation and proteolysis frequently drive proteins to irreversible unfolded states. The Lumry-Eyring model (Lumry and Eyring 1954) is frequently used to describe protein conditions where irreversible denaturation occurs in a relevant time scale:

$$\mathbf{N} \leftrightarrow \mathbf{U} \rightarrow \mathbf{F} \qquad \textbf{[I.2]}$$

where N, U, and F represent, respectively, native, unfolded, and final irreversible states. Therefore, under these conditions the rate of function loss (*i.e.*, the rate of N to U conversion) becomes biologically very significant. In physiological conditions, a protein is kinetically stable when its free energy barrier is sufficiently high to keep the unfolding rate low in relation to the time-frame necessary for function execution. An interesting example of a protein under kinetic control is the α-lytic protease which is synthesized as a pro-enzyme that needs to be cleaved to yield the active form (Baker et al. 1992; Baker and Agard 1994; Jaswal et al. 2002).The cleaved region is fundamental for correct folding, therefore, after being removed the protein is no longer able to refold to the active form. In physiological conditions, the

active state is thermodynamically less stable than the inactive state. However, the high energy barrier keeps the protein in the active state for several months (half-life of about one year). Hence, the relevant parameter for this protein is the kinetic stability. Another example showing the importance of kinetic stability is the comparison of citrate synthase from organisms with different optimal growth temperatures, in the range of 30°C to 100°C. In this case, the temperature necessary for the protein to display a half-life of 8 minutes progressively increases with the optimal growth temperature of the organism: from 45°C for the synthase of the most mesophilic organism to 100°C for the synthase of the most thermophilic one (Bell et al. 2002a).

Forces that drive proteins to the folded state

Proteins are in a constant unfolding/refolding balance, but under physiological conditions the fraction of molecules populating the unfolded state is very small. Therefore, the contributions of the forces that determine the structure of the native and unfolded states have to be considered in order to understand the net result to the stability of the protein.

There are different types of forces which can combine or cancel out resulting in a particular degree of protein stability:

i) Under the hydrophobic collapse hypothesis (Vidugiris et al. 1995; Agashe et al. 1995), protein folding is driven mainly by hydrophobic interactions which, by removing non-polar protein units from solvent contact, maximize water entropy thus lowering the total free energy. The energetic contribution of hydrophobic collapse has been estimated to be around 1 kcal/mol per methylene group that is removed from water contact (Serrano et al. 1992; Pace 1992). It is considered that only after this initial protein collapse do other forces begin to contribute to protein folding and thus to its stability, namely hydrogen bonding, electrostatics and van der Waals;

ii) Backbone hydrogen bonds are particularly important in determining the development of secondary structural elements while side-chain hydrogen bonds appear to have a greater role in the formation of the tertiary structural arrangement. Hydrogen bonding normally occurs between atom pairs located within 2.5 Å (shorter than the standard van der Waals distance) and their energy is estimated to vary between 3 and 14 kcal/mol. However, proteins are not hundreds of kcal/mol stable, which is explained by the fact that in the unfolded state most of the potential hydrogen bonding partners are stabilized with hydrogen bonds to water, thus contributing to the stability of this state (McDonald and Thornton 1994). However, breaking of these protein-water hydrogen bonds upon folding contributes to an increase in water entropy. The fine balance between the enthalpic and entropic terms of hydrogen bonding determines whether this type of interaction has a positive contribution to protein stability or not (Rose et al. 2006; Pace et al. 1996). Although the actual contribution of hydrogen bonds to protein stability is not trivial to estimate, the value of 0.6 kcal/mol per hydrogen bond has been accepted as a good approximation (Fersht 1987; Pace et al. 1996). In the case of *Candida methylica* formate dehydrogenase, hydrogen bonding was estimated to stabilize the protein by 4 kcal/mol (Karaguler et al. 2004) in fair agreement with the approximation mentioned above. More recently, it has been shown that hydrogen bonds established within the protein hydrophobic core have a greater contribution to protein stability than those exposed to the more polar water environment, since they stabilize this hydrophobic core (Deechongkit et al. 2004);

iii) van der Waals forces, or London dispersion forces, are weak and short ranged but their cumulative effect may produce a significant contribution to protein stability (Nölting 1999; Deechongkit et al. 2004). These interactions are produced between two instantaneous dipoles and have an attractive and a

repulsive term. As a curiosity we should mention that van der Waals forces have been put forward as a possible explanation for the ability of certain small animals (*e.g.*, the lizard Gecko) to climb on sheer surfaces (Autumn et al. 2002; Deechongkit et al. 2004). Unlike hydrogen bonds, van der Waals interactions have no significant entropic component and rely mainly on the enthalpic term (Cooper 1999);

iv) Electrostatic forces can stabilize proteins by up to 1 kcal/mol when located in the interior of the protein and 0.5 kcal/mol if located at the surface. This is mostly explained by the decrease of the interaction intensity with the dielectric constant of the medium. In general water has a dielectric constant 40-fold greater than that observed for the interior of proteins (Murphy 1995; Mozhaev 1993), although the ionic content of water can significantly affect this relation. Importantly, long range ionic interactions imply a significant entropic penalty to the folded state, however, subsequent ionic interactions may not be subjected to the same penalty. Therefore, while a single ionic interaction can stabilize or destabilize the native state, networks of ionic interactions are generally strong promoters of the native state (Lebbink et al. 1999).

Attempts to estimate the contribution of the various forces to the net stability of the folded state appear to match the order of magnitude of those contributions but fail in attaining a satisfactory result for the true free energy value (Pace et al. 1996; Casadio et al. 1995). Table I-1 reports an example of this difficult exercise with the protein RNase T1.

Table I-1. Stabilizing and destabilizing interactions acting on the folding/unfolding equilibrium of RNase T1 (adapted from Pace et al. 1996).

Interaction type	Contribution to free energy variation
Destabilizing	**kcal/mol**
Conformational entropy	- 177
Peptide groups buried	- 81
Polar groups buried	- 28
TOTAL (Dest.) =	**- 286**
Stabilizing	**kcal/mol**
Histidine ionization	+ 4
Disulfide bonds	+ 7
Hydrophobic groups buried	+ 94
Hydrogen bonding	+ 166
TOTAL (Stab.) =	**+ 271**
ΔG predicted (Stab. – Dest.) =	- 15 kcal/mol
ΔG experimental =	+ 9 kcal/mol

A glimpse of the unfolded state

The unfolded state of proteins has been regarded for many years as the "black sheep" of the protein kingdom. It is misbehaved, complex, structurally ill-defined, difficult to study and most often an unwanted experimental reality. However, its importance in protein folding and stability becomes apparent from the previous section and has long been recognized, as shown by the several reports dedicated to its study in the early days of the protein folding field (Baldwin 1986; Miller and Goebel 1968; Tanford 1968; Tanford 1970). More recently, new analytical methodologies supported by great instrumentation advances, namely in small angle X-ray scattering (SAXS), NMR, and Fluorescence resonance energy transfer (FRET), allowed probing the unfolded state with unprecedented detail, dramatically changing our view

of it. From the initial assumptions that the unfolded state was a random coil, an ensemble of extended conformations with basically no energy barriers between interconverting states, we have now strong evidence for the existence of significant structural features populating the unfolded protein state (Robin et al. 2002; Ratcliff and Marqusee 2010; Pletneva et al. 2005; Pan et al. 1995). A classic example of residual structure in the unfolded state is given by ribonuclease A, with SAXS experiments showing that even under strongly denaturing conditions the gyration radius of this protein remains half the value predicted for a random coil (Sosnick and Trewhella 1992). However, these features generally comprise a small fraction of the amino acid sequence, limiting the utility of techniques based on overall secondary structure content. In respect to NMR, the availability of very strong magnetic fields, more sensitive cryo-probes, and developments in methodologies like relaxation dispersion (Korzhnev et al. 2004), residual dipolar couplings (Nodet et al. 2009), long range NOEs (Crowhurst and Forman-Kay 2003), paramagnetic relaxation enhancement (Gillespie and Shortle 1997; Lietzow et al. 2002) and protein selective labelling have made the unfolded state accessible in much greater detail. Most importantly, NMR allows unfolded protein states to be probed under non-denaturing conditions (Korzhnev et al. 2004; Tollinger et al. 2001). FRET kinetics has been shown to be very useful in identifying short and long-range interactions as well as in characterizing the conformational heterogeneity of denatured proteins (Pletneva et al. 2005), with the advantage of using less bulky probes (when compared to paramagnetic relaxation enhancement). Importantly, a recent report has shown it to be possible to introduce two probes (ReAsH and FlAsH) orthogonally in a single protein (Zurn et al. 2010); as these probes are capable of crossing cell membranes, FRET measurements can be considered for *in vivo* experiments (Enninga et al. 2005). Finally, computational studies assume a major role in

the visualization of the consequences of experimental constraints and in checking consistency (Bernado et al. 2007; Cho et al. 2007; Dedmon et al. 2005).

The presence of structural features in the unfolded state of several proteins implies a restriction of the conformational space. This reasoning suggests an unfolded state described by a dynamic ensemble of multiple conformations but where not all conformations are equally allowed (Shortle 1996; Barbar et al. 1996). Thus the free energy of the unfolded state should be represented by a hilly plateau which is separated from the native state by a high energy barrier corresponding to the transition state (Finkelstein and Ptitsyn 2002). One important implication arising from the unfolded state not being a random coil in a flat energy landscape is that changes in the protein sequence (*i.e.*, mutations) may affect both the folded and unfolded conformational ensembles of a protein.

Despite great advances in the characterization of the unfolded state, it remains an intrinsically difficult object of study for which we are only allowed to have a glimpse. Still, knowledge on the unfolded state is expected to increase substantially in years to come, in particular due to the discovery of the important role of intrinsically unfolded proteins (commonly named intrinsically disordered proteins – IDPs) in central biological processes such as transcriptional regulation, post-translational modification, signalling, and protein aggregation (Fuxreiter et al. 2008; Turoverov et al. 2010). It is now accepted that IDPs are not outliers; indeed, they are major players in complex biological functions and probably the key factors in the development of highly complex organisms. Not surprisingly many IDPs are strongly linked to complex debilitating and sometimes life threatening diseases including diabetes, Alzheimer's, Parkinson's, and cancer (Iakoucheva et al. 2002; Uversky et al. 2008).

Like the unfolded state of many globular proteins, IDPs exhibit residual structure or topological features in the conformational ensemble. In many cases, these features appear to be fundamental to the recognition/binding process of IDPs to the respective binding partners (Wright and Dyson 2009; Ganguly and Chen 2011; Cheng et al. 2007). The interest in the IDPs has increased tremendously due to possible implications in human related diseases. Consequently the interest in the characterization of highly disordered protein structures has also grown exponentially which will surely benefit not only the IDP field but also the field of globular proteins.

PROTEIN STABILIZATION BY COMPATIBLE SOLUTES

Compatible solutes from hyper/thermophiles

One of the common strategies adopted by microorganisms to cope with environmental stress, namely high osmotic pressure or heat, is the accumulation of low-molecular mass organic compounds in the intracellular space. These compounds are usually named "compatible solutes" (Brown and Simpson 1972; Brown 1976) due to their non-interference or "compatibility" with cell metabolism. However, the designation "osmolytes" is frequently used to highlight their major role in balancing the external osmotic pressure. In general, we favour the designation "compatible solutes" since it applies to any type of stress. The word "thermolyte" has been proposed (Santos et al. 2011) to specify the organic solutes that accumulate in response to supra-optimal temperature, and supposedly are involved in thermoprotection. In this thesis, the designations osmolyte, compatible solute, solutes and co-solutes will be used arbitrarily to mean organic compounds that usually enhance protein stability.

Compatible solutes belong to a restricted class of chemical compounds, namely amino acids and derivatives, sugars and polyols, but some are associated with microorganisms adapted to specific environments. For example, di-*myo*-inositol phosphate, diglycerol phosphate and mannosylglycerate are preferentially accumulated by heat adapted organisms (Santos and da Costa 2001; Santos and da Costa 2002) (Figure I-1). Also, the patterns of solute accumulation frequently depend on the type of imposed stress. For example, in a variant of *Archaeoglobus fulgidus*, diglycerol phosphate is the major solute when cells are subjected to salt stress, whereas di-*myo*-inositol phosphate is the predominant solute under heat stress (Gonçalves et al. 2003). Most interesting, compatible solutes found in hyper/thermophilic organisms are often negatively charged, while mesophilic organisms tend to accumulate neutral or zwitterionic compounds (Roberts 2005; da Costa et al. 1998). This intriguing difference is certainly not a coincidence and may have several origins. It has been suggested that negatively charged compatible solutes may serve to balance the presence of potassium ions that the cell takes up from the medium as a first response to osmotic stress (Roberts 2005). For example, in the thermophilic *Methanococcus thermolithotrophicus*, the accumulation of glutamate follows very closely the intracellular levels of potassium ion (Robertson et al. 1992). Up to 1 M of external NaCl, these cells respond to the osmotic stress by taking up potassium ions from the medium with glutamate isomers being the main accumulated solutes. At 1.7 M of external NaCl, the internal concentrations of potassium and glutamate decrease to half the concentration observed at 1.3 M of NaCl. In turn, Nε-acetyl-β-lysine (zwitterionic) replaces glutamate as the major accumulated solute thus supporting the notion that the net charge of solutes may serve to counter balance the accumulation of potassium cations. On the other hand, *in vitro* experiments, have shown that charged

compatible solutes of hyper/thermophiles reduce protein aggregation and unfolding to a greater extent and at lower concentrations than neutral solutes (Ramos et al. 1997; Faria et al. 2008; Borges et al. 2002). These properties may be particularly relevant for hyper/thermophiles, since aggregation and unfolding phenomena should be more likely at the higher temperatures that characterize the environments of these organisms. In apparent support of this view is the observation that *Rhodotermus marinus* cells accumulate preferentially mannosylglycerate (negatively charged) under heat stress (77°C), while mannosylglyceramide (neutral) is the main solute accumulated under osmotic stress (5% NaCl) (Silva et al. 1999). In fact, mannosylglycerate is, in general, a much better protein stabilizer than mannosylglyceramide (Ramos et al. 1997; Faria et al. 2008; Borges et al. 2002).

The preferential accumulation of negatively charged compounds by hyper/thermophiles remains intriguing, but it is tempting to speculate that these solutes were selected through evolution as the best thermostabilizers of macromolecules and other cellular components.

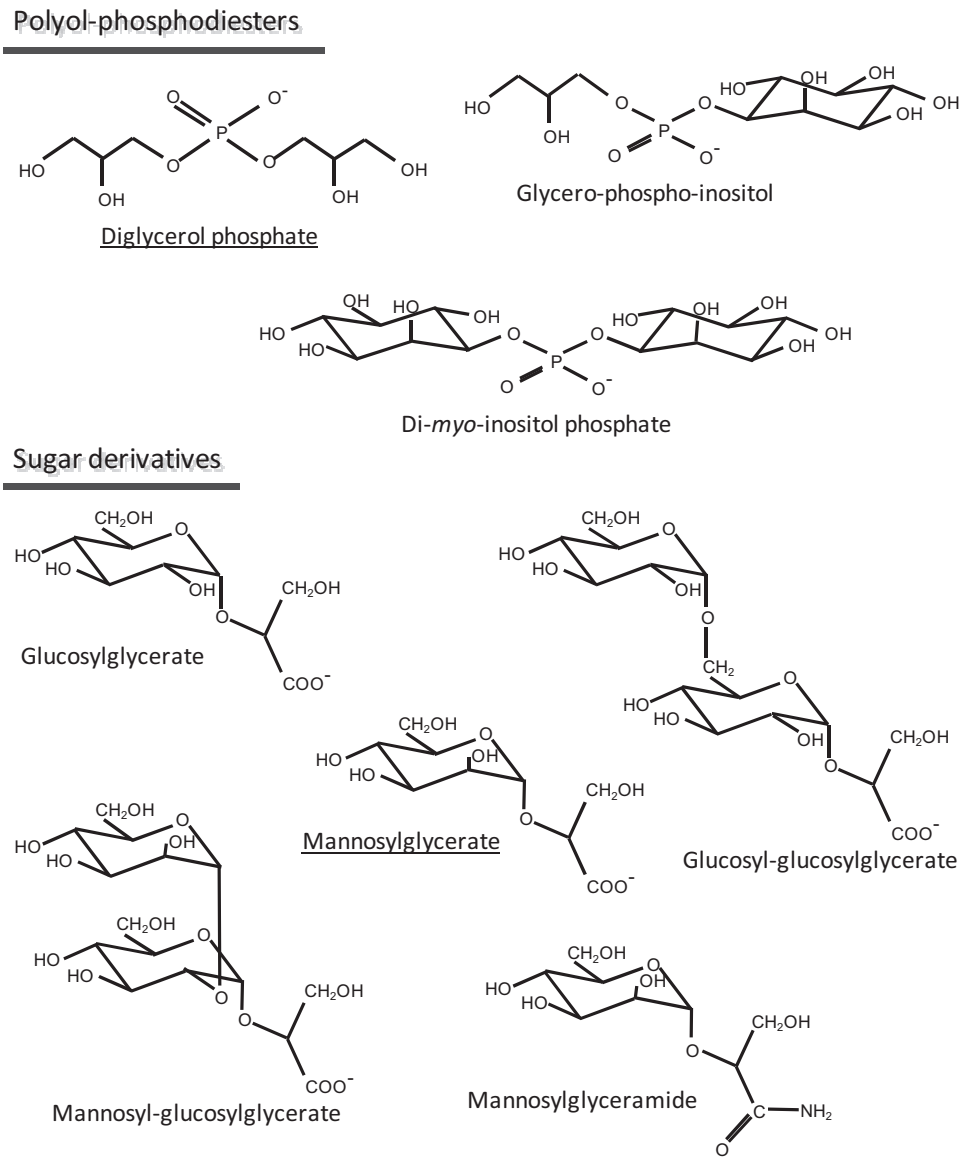


Figure I-1. Chemical structures of compatible solutes highly restricted to (hyper)thermophilic organisms.

Control of the folding/unfolding equilibrium by solutes

The study of the principles that govern protein stabilization by compatible solutes needs to address the several components in the system as well as their combined effect. Special attention has to be given to the interaction of the solute-water system with both extremes of the protein energy landscape, as well as to the interaction of the solute with the water structure (Davis-Searles et al. 2001; Ramos and Baldwin 2002; Timasheff 1993). The interaction of a solute with the protein surface can result in solute exclusion from the contact interface or solute binding to the protein surface. Solute binding can be non-specific, resulting in a higher concentration of solute molecules in the solvent-protein interface than in the bulk, or specific in which regions of the protein with certain chemical properties favour the interaction. On the other hand, solute exclusion occurs mainly due to: i) steric effects, when the solute molecules are much bigger than water as in the case of the interaction of polyethylene glycol with β-lactoglobulin (Arakawa and Timasheff 1985); ii) increase of water surface tension, caused by solute interference with the water structure, and iii) solvophobic effects, due to unfavourable interactions with non-polar regions of the protein, as in the case of glycerol (Gekko and Timasheff 1981). Whichever the type of interaction established by stabilizing or destabilizing solutes with the protein or with the water, they will affect both the folded and unfolded states of the protein but with different energetic outcomes. Therefore, the magnitude of protein stabilization is a function of the net effect of the solute with the two states of the protein. However, these interactions are weak, requiring great accuracy to assess their energetic contributions.

The search for a general thermodynamic model that can explain the stabilizing interactions and the magnitude of such contributions has been the

focus of many research studies since the middle of last century; some of the most relevant ones are described below.

In the early 1960's, Nozaki and Tanford addressed the solubility of amino acids and peptides in a variety of solutions namely water, urea, ethylene glycol and guanidine hydrochloride (Nozaki and Tanford 1963; Nozaki and Tanford 1965; Nozaki and Tanford 1970; Gekko and Timasheff 1981). They found that urea solutions were able to accommodate non-polar amino acid side-chains and peptide groups better than water. As the unfolded state exposes more of these units to the solvent, its free energy decreases more than that of the folded state, thus inducing unfolding. Nozaki and Tanford considered that the energetic contribution of each amino acid in a protein molecule to the folding/unfolding equilibrium was additive. Thus, they proposed that the free energy of transfer of the protein from water to another solution could be calculated based on the free energy of transfer of each amino acid, using the following equation:

$$\Delta F_u = \Delta F^0_{u,\mathrm{H_2O}} + \sum_i \alpha_i n_i^0 \Delta f_{t,i} + \delta \Delta W \qquad \textbf{[I.3]}$$

in which ΔF_i is the free energy change of protein unfolding, α_i is the average fractional change in solvent exposure when group *i* goes from the folded to the unfolded state, $\Delta f_{t,i}$ is the free energy of transfer of a group *i* when the protein is transferred from water to another solvent, n_i^0 is the number of such units present in the protein, and $\delta\Delta W$ the change in the contribution of electrostatic interactions to the free energy of unfolding (Nozaki and Tanford 1964; Nozaki and Tanford 1970). This is called the transfer free energy model and can be represented by the thermodynamic cycle depicted in Figure I-2.

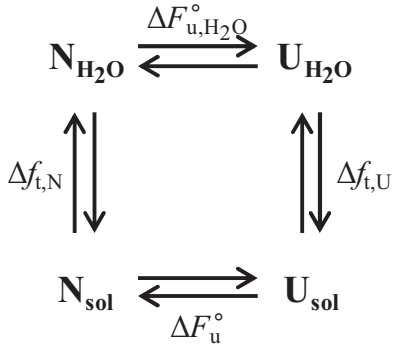


Figure I-2. Thermodynamic representation of the transfer free energy model. Vertical arrows represent the transfer of the native (N) or the unfolded (U) protein from water to another solution, while horizontal arrows represent the unfolding of the protein in pure water (top) or in another solution (bottom).

Later, Timasheff and co-workers performed densitometry studies on peptide solutions containing several co-solutes, namely sucrose and glycerol. Based on the Wyman relation (Wyman 1964) they found negative binding to the protein surface for a number of stabilizing agents (Arakawa and Timasheff 1985). It was only after the studies of John Schellman on multi-component thermodynamics (protein – water – solute) that researchers realized how positive equilibrium constants could lead to negative stoichiometries (Schellman 1987b; Schellman 1993). The negative binding calculations suggested that the solute was being preferentially excluded from the protein-water interface giving rise to the preferential hydration theory (Timasheff 1993; Arakawa and Timasheff 1985). According to this theory, solute exclusion is a thermodynamically unfavourable phenomenon that gains expression upon protein unfolding due to the expansion of the water-protein interface (Figure I-3).

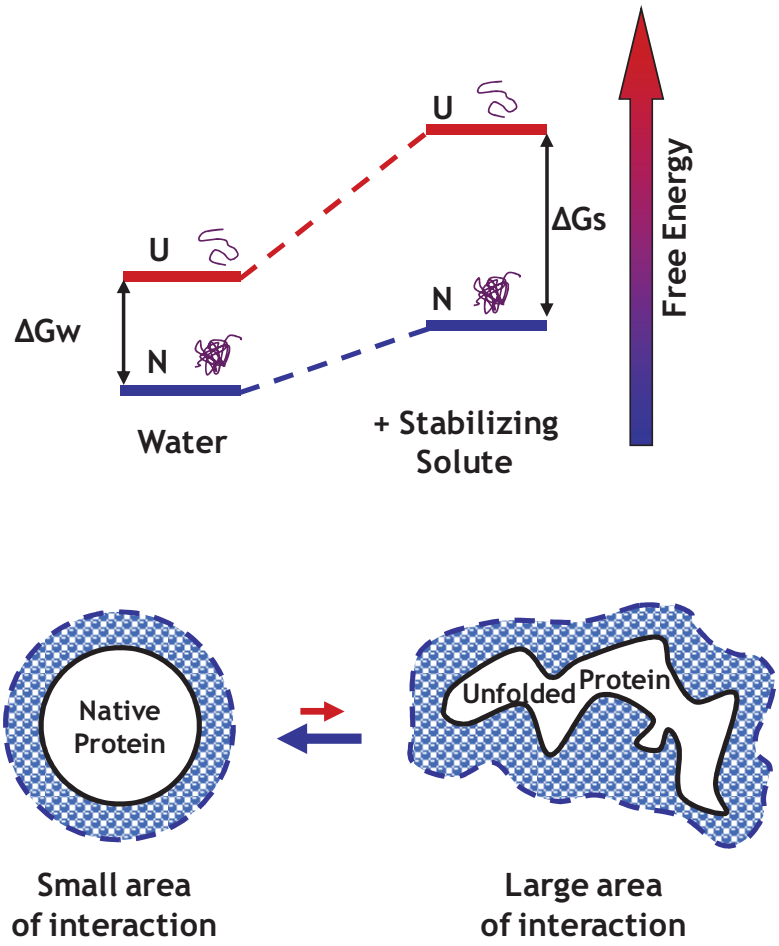


Figure I-3. Representation of the thermodynamic effect of preferential hydration in the presence of a stabilizing solute. Preferential hydration by solute exclusion is a thermodynamically unfavourable process meaning that the energy in the system increases. However, because the volume of exclusion increases upon unfolding due to the larger area of interaction, the equilibrium is shifted to the folded conformation, thus stabilizing it.

More recently Bolen and colleagues revisited Tanford's transfer free energy model and went further both in the interpretation of the type of solute-protein interactions and in the quantitative determination of transfer free energies of model peptides from water to osmolyte solutions. They realized that the

widely accepted view of surface tension as the source of solute exclusion was useful in explaining the overall thermodynamics of protein stabilization in physical terms but lacked the insight about the chemical origins of the unfavourable interactions of stabilizing agents with the protein. Then they picked up on Tanford's transfer free energy model and measured the interaction between stabilizing solutes and exposed chemical groups of native and unfolded proteins (Liu and Bolen 1995). In these studies they found that the major driving force for protein stabilization was the highly unfavorable interaction of the backbone with the solute which drives the equilibrium to the native state in order to minimize backbone exposure and thus unfavourable interactions. As consequence of this unfavourable interaction the denatured ensemble becomes contracted and hydrophobic packing increases, further contributing to protein folding (Qu et al. 1998). The peptide backbone is also responsible for protein unfolding in the presence of urea since it establishes favourable interactions with this compound. On the other hand, side-chains show a variable and weak contribution to the thermodynamic stability either in the presence of either stabilizing or destabilizing compounds.

Bolen and colleagues improved the transfer free energy model by showing that it was possible to overcome several of its limitations, namely: i) proper demonstration that transfer free energies of the different protein groups were additive, ii) dependence of the transfer free energies on the chemical model used to represent the peptide backbone and on the concentration unit used to describe solubility (*i.e.*, molarity, molality or mole fraction), and iii) lack of knowledge about the real contribution of activity coefficients to the transfer free energy measurements (Auton and Bolen 2004; Hu et al. 2010). The validity of this model has been subjected to a stringent test in a recent publication where 46 proteins and 9 solutes showed a 1:1 correspondence between predicted and measured thermodynamic stabilization – although the

large majority of the predictions were done for the presence of urea (Auton et al. 2011). Nevertheless, care must be taken with the model of the denatured state since, as mentioned before, the denatured state can vary from simple random coil, to more complex configurations with different degrees of residual structure.

Despite this apparent success, all of these attempts to explain protein stabilization by compatible solutes have a potential source for uncounted contributions: the role of the folding/unfolding pathway and of the respective kinetic intermediates.

The importance of the solute charge in protein stabilization

Most studies that address the mechanism of protein stabilization by compatible solutes have been performed using neutral or zwitterionic compounds that often require high concentrations (usually in the molar range) to produce a measurable impact on protein stability. However, charged compatible solutes typical of marine hyper/thermophilic organisms are usually found to significantly stabilize proteins at much lower concentrations namely between 0.10 and 0.50 M (Lamosa et al. 2000; Faria et al. 2008; Borges et al. 2002). In fact, the net charge seems to be decisive for the superior stabilization rendered by these solutes. For example, the effects of several charged solutes namely mannosylglycerate, mannosyl-lactate, mannosylglycolate, glucosylglycerate, and diglycerol phosphate were shown to render greater thermal-protection to three enzymes (malate dehydrogenase, staphylococcal nuclease and lysozyme) when compared to neutral solutes such as mannosylglyceramide, mannosylglycerol, glycerol, ectoine and hydroxyectoine (Faria et al. 2008). Another study concluded that the stabilizing effect of several carboxylic acids on four proteins increased with the number of carboxylic groups in the salt, underlining the contribution of the

charged groups to protein stabilization (Kaushik and Bhat 1999). The notable importance of the net charge coupled with the much lower concentrations at which these solutes stabilize proteins, suggests that the mechanisms of stabilization may differ from those proposed for common non-charged solutes. As described above, the stabilizing effect of solutes has been explained as arising from the interactions of the solute with the protein and with the water structure which ultimately result in different degrees of solute exclusion from the protein surface. Interestingly, arginine, a charged molecule which can stabilize proteins and suppress aggregation (Vedadi et al. 2006; Umetsu et al. 2003; Bell et al. 2002b), binds to bovine serum albumin up to 0.5 M, but becomes progressively more excluded from the protein surface at higher concentrations (Arakawa et al. 2007). However, it was shown that arginine reduces protein unfolding and aggregation at concentrations ranging from 0.1 to 1 M (Bell et al. 2002b; Kudou et al. 2003; Umetsu et al. 2003). Therefore, solute exclusion may not be the only way by which arginine induces protein stabilization. It has been proposed that arginine molecules may form clusters in solution which can specifically interact with the protein surface reducing solvent access to the hydrophobic regions of the protein (Das et al. 2007). Important clues may also be obtained from the study of protein stabilization by inorganic ions. The stabilization by inorganic salts has been interpreted as resulting from the balance of surface tension increase and weak interactions between the ions and the protein groups (Arakawa and Timasheff 1984; Tadeo et al. 2009). Therefore, we should be aware that charged compatible solutes may exhibit different patterns of interactions with the protein and water. These interactions may be difficult to rationalize under the scope of the current models of protein stabilization.

The molecular mechanisms involved in protein stabilization have long been sought and, despite great advances in its understanding, the discussion

remains far from closed as many observations still escape a clear theoretical basis.

PROTEIN DYNAMICS AND ITS RELATIONSHIP WITH PROTEIN STABILITY

Proteins are highly dynamic entities, not only in the unfolded state but also in the native state. In fact, dynamics is virtually as important to protein function as the protein structure itself. Ligand recognition, binding and catalysis are intimately related to protein internal mobility (Kay et al. 1998; Kay 2005; Larsen et al. 1997). For example, the X-ray structure of haemoglobin shows an apparent inaccessible haem centre, requiring conformational rearrangements of the side-chains in order for the oxygen to be able to reach it (Perutz and Mathews 1966). Also, the right level of mobility seems to be the reason why most enzymes work better at the edge of their thermal stabilities. This has brought about the idea of "corresponding states", according to which homologous proteins with different stabilities show similar flexibility and activity at their optimal temperatures (Daniel and Danson 2001; Jaenicke et al. 1996).

Protein motions cover a wide range of time scales and amplitudes and together they give shape to the conformational ensemble that characterizes protein structure. NMR spectroscopy is the leading tool for studying protein dynamics due to its versatility and precision. With this approach, atomic level resolution becomes available for the study of a wide range of protein motions, from sub-nanosecond bond fluctuations to millisecond translations of atom groups to even slower opening motions of whole structural elements. Although NMR has been used for a long time to study protein dynamics (Richarz et al. 1980; Allerhand et al. 1971) it was only after the development

of the multidimensional spectra methodology (proposed by Jean Jeener in 1971, but largely developed by Richard Ernst, the 1991 Nobel Laureate in Chemistry) and efficient isotopic protein labelling protocols that NMR became a powerful tool in the study of protein dynamics. A summary of the various NMR measurements and respective time scales probed is shown in Figure I-4.

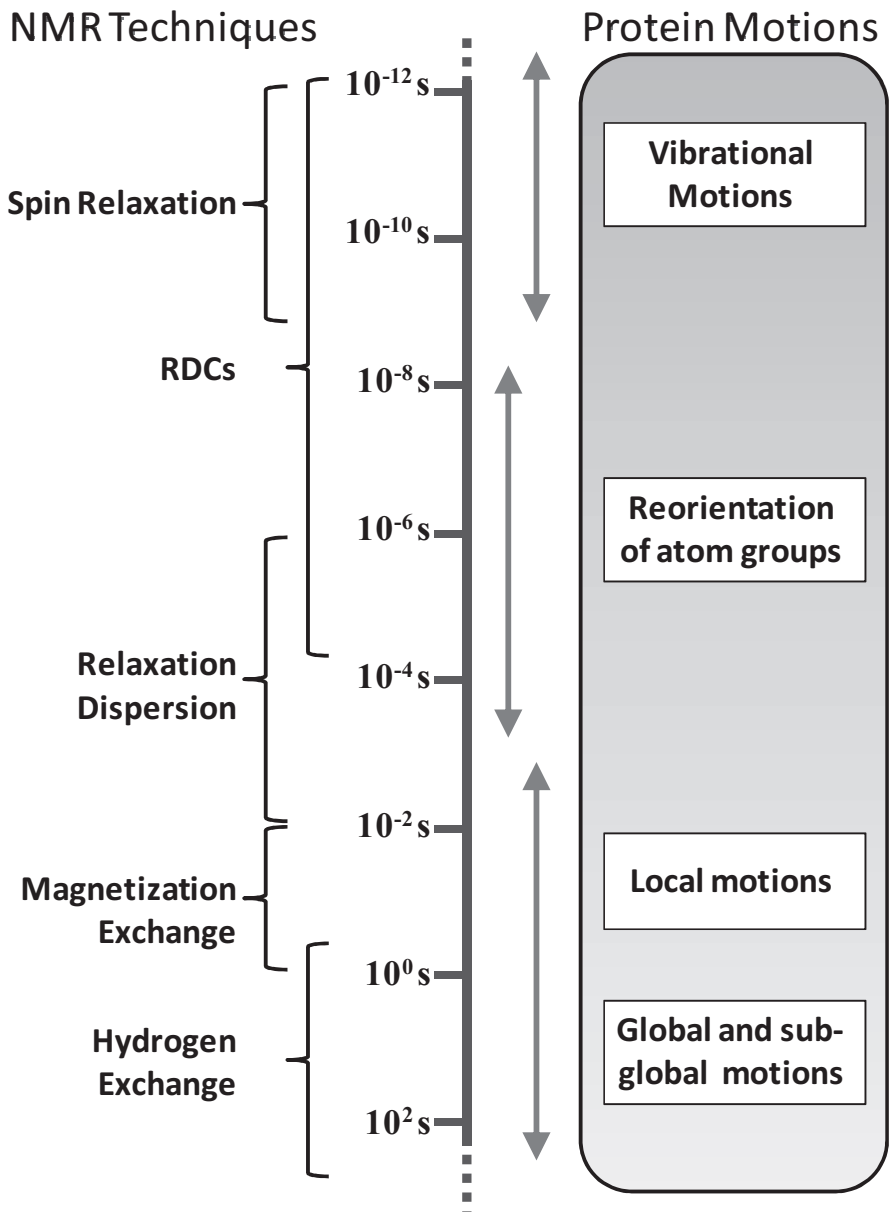


Figure I-4. Overview of the protein motions that characterize the conformational ensemble and depiction of NMR techniques used to probe each time scale.

Protein motions imply energy changes, therefore large global or nearly global motions are less frequent than small amplitude local motions which involve smaller energy variations. Additionally, protein motions may facilitate solvent access to the hydrophobic interior thus interfering with proper packing. So, if a protein is less flexible, then one might expect an increased stability. This is the case of two homologous 3-isopropylmalate dehydrogenases from the thermophilic *Thermus thermophilus* HB8 and from the mesophilic *Escherichia coli* where hydrogen exchange studies revealed that at room temperature the thermophilic protein was more rigid (Zavodszky et al. 1998). Another study with a group of *p*-nitrobenzyl esterase variants obtained by directed evolution, revealed that increased protein stability was often accompanied by decreased flexibility (Gershenson et al. 2000). However, does it follow that increased thermal stability needs to be accompanied by decreased flexibility of the protein native state? In other words, is there a correlation between protein stability and protein dynamics? This intuitively attractive notion was thought to be the general trend for some time (Vihinen 1987; Varley and Pain 1991; Tang and Dill 1998). However, with the development of new techniques and more protein models, it became clear that the matter is not as simple as initially considered, with several studies challenging this trend (Aguilar et al. 1997; Hernandez et al. 2000; LeMaster et al. 2005). For example, swapping the multi-turn domain of the mesophilic *Clostridium pasteurianum* rubredoxin with that of the hyperthermophilic *Pyrococcus furiosus* rubredoxin results in a protein hybrid with a melting temperature (corresponding to the reversible thermal unfolding) 12°C higher than the that of mesophilic rubredoxin. However, the backbone flexibility of the more stable hybrid in the sub-second time scale is equal to or greater than that of the mesophilic rubredoxin (LeMaster et al. 2005). Another study showed that the ribonuclease from the thermophilic *Thermus thermophilus* was more mobile in the microsecond to

millisecond time scale than the ribonuclease from the mesophilic *Escherichia coli* (Butterwick et al. 2004). On the other hand, no general trend was found for the degree of mobility at the sub-nanosecond time scale of both ribonuclease homologues: while two β-sheets and a glycine rich region are more rigid in the thermophilic ribonuclease, the substrate binding region and the C-terminal tail are more flexible. In fact, the study of the relationship between protein dynamics and protein stability is complicated by several issues: precise quantification of changes in dynamics, relative importance of the different time scales and amplitudes, and relative importance of different regions of the protein (as clearly illustrated by the ribonuclease study just mentioned). All these issues have, so far, prevented a generalized consensus on the relationship between protein dynamics and protein stability, despite the fact that an inverse relationship has been, in fact, demonstrated for some proteins.

EXPLORING PROTEIN MOTIONS BY NMR

NMR spectroscopy has clearly been acknowledged as the biophysical technique which can provide the most comprehensive and reliable information about protein motions. Nearly all heavy atom positions can be used to probe the amplitude and time scales of conformational dynamics. Methodological developments have made such progress that it is now possible to probe motions associated with sparsely populated native states, frequently called "invisible states" or "excited native states" (Korzhnev and Kay 2008). Furthermore, NMR studies can be conducted under a variety of sample conditions that greatly expand the analysis possibilities: pH, temperature, and chemical composition (*e.g.*, salt content, co-solute content, buffer, etc.) of the solution can be changed without interfering with the probing of protein

dynamics. NMR is therefore the technique of choice when considering a comprehensive study of the effect of stabilizing solutes on the whole range of motions experienced by proteins. Specific isotopic labelling of the proteins provides a useful means to avoid interference of the solute signals in the NMR spectra used to measure spin relaxation of the protein nuclei.

However, there are also limitations to the use of NMR in the study of protein motions, namely: i) protein size, resonance assignment becomes increasingly difficult and resolution becomes poorer, ii) signal detection, NMR is a low sensitivity technique requiring relatively large amounts of biological material (~0.2 mM in a 400 µL volume sample) when compared to other biophysical techniques, and iii) interpretation of relaxation data in terms of dynamic behaviour can be complex and may, sometimes, be obscured by cross interference or concurrent relaxation phenomena that do not report purely on internal protein motions. These limitations are constantly being pushed forward by technological advances such as the development of higher field magnets, more sensitive probes (*e.g.,* cryo-probes), new or improved pulse sequences to cope with protein size limitations, and improved isotopic labelling schemes to reduce relaxation interference phenomena.

Nevertheless, comparison with other biophysical methods is advisable not only to cope with NMR's own limitations but also to cross validate the extracted information. Alternative techniques that are frequently employed in the study of protein dynamics include fluorescence spectroscopy, molecular dynamics simulations (MD), incoherent neutron scattering and X-ray diffraction:

- Fluorescence spectroscopy often uses the aromatic tryptophan and tyrosine residues of the protein as intrinsic probes to obtain information on overall rotational diffusion and internal motions. It is also possible to attach other fluorescent probes to the protein, but this

procedure is generally avoided due to the unpredictable effects they may have in the protein motions. In any case, the time scales probed are on the order of the fluorescence life time of the fluorophore in use. Quantitative information is complicated by multi-exponential fluorescence decays, therefore most studies are conducted in proteins containing a single fluorophore, usually tryptophan, but a few notable examples using tyrosine residues have been reported (Noronha et al. 2009; Noronha et al. 2010). Tryptophan fluorescence can be measured without interference from tyrosine residues by keeping the excitation wavelength above 290 nm. However, fluorescence studies of protein motions in the presence of stabilizing solutes is limited to compounds with absorbance and fluorescence properties that do not obscure the signal of the protein;

- MD simulations have become a valuable method for the visualization of the dynamic information obtained from NMR analysis. They can also provide atomic level resolution but computational demands limit the analysis to the nanosecond time scale;
- Incoherent neutron scattering is sensitive simultaneously to the amplitudes and frequencies of the protein motions. The interaction of the neutrons with the protein is made via the sample nuclei therefore, with proper isotopic labelling this approach provides site-specific dynamic information. However, this technique requires large amounts of sample material (~100 mg);
- Finally, X-ray crystallography can be used to obtain thermal disorder information of the protein atoms. B-factors, or Debye-Waller factors, are sensitive to atom displacements due to thermal motion and with the proper experimental analysis can provide qualitative dynamic information in agreement with the generalized order parameters

obtained from NMR. However, no information can be extracted about the time scales of the motions. Moreover, the motion constraints associated with protein packing in the crystal cannot be excluded.

In the NMR area, there are specific methodologies to address protein motions of different time scales and amplitudes: i) Spin relaxation measurements provide generalized order parameters of bond motions precessing in the sub-nanosecond time scale, ii) relaxation dispersion measurements are well suited to study conformational exchange on the microsecond to millisecond time scale, iii) magnetization exchange experiments report on the high millisecond time scale conformational exchange, and iv) Hydrogen/Deuterium exchange, one of the earliest methods used to show the dynamic nature of proteins, reports on local and global conformational fluctuations in the time scale of millisecond and beyond.

Picosecond to nanosecond bond motions

Spin relaxation measurement (usually of ^{15}N, ^{13}C and ^{2}H) is a well established technique in the study of protein motions up to a few nanoseconds. Backbone motions are usually studied by measuring ^{15}N relaxation rates while side-chains are probed using ^{13}C or ^{2}H relaxation data. Relaxation is the return of the spin populations to their equilibrium values after an energy perturbation. In liquids, spin relaxation is caused by fluctuating local fields which include chemical shielding anisotropy (CSA), dipole-dipole interactions, quadrupolar interactions and sometimes paramagnetic interactions (Abragram 1961). The contribution of these fluctuating fields to spin relaxation depends on how fast the molecules in a sample are moving hence the reason for relaxation measurements encoding dynamic information. In ^{15}N and ^{13}C relaxation, the dominant relaxation mechanisms are dipole-dipole interactions and CSA while in ^{2}H, quadrupolar relaxation dominates. With the proper pulse sequences

(Skelton et al. 1993; Nicholson et al. 1992; Kay et al. 1989; Farrow et al. 1994) and isotopic labelling it is possible to obtain pure relaxation rates, that is, to measure relaxation rates due to a single relaxation mechanism (*e.g.,* dipole-dipole). It is very important to extract accurate spin-lattice (T_1) and spin-spin (T_2) relaxation times, the parameters that characterize spin relaxation. Spin-lattice relaxation results from spin interactions with the lattice such as solvent molecules or other nuclei within the molecule and induces spin transitions between energy levels, while spin-spin relaxation involves energy exchange between spins and does not alter the overall population distribution between energy levels. Importantly, both processes contribute to the decay of the NMR signal.

In pure relaxation conditions, the rate constants and steady state heteronuclear NOE values can be calculated by the following expressions (Abragram 1961):

$$\boldsymbol{T_1^{-1}} = (nd^2/4) \times [J(\omega_X - \omega_A) + 3J(\omega_A) + 6J(\omega_X + \omega_A)] + c^2 J(\omega_A) \quad \textbf{[I.4]}$$

$$\boldsymbol{T_2^{-1}} = (nd^2/8) \times [4J(0) + J(\omega_X - \omega_A) + 3J(\omega_A) + 6J(\omega_X + \omega_A)] + 6J(\omega_X)] + c^2/6 \times [4J(0) + J(\omega_A)] \quad \textbf{[I.5]}$$

$$\mathbf{NOE} = 1 + nd^2\gamma_X/4\gamma_A \times [6J(\omega_X + \omega_A) - J(\omega_X - \omega_A)]/\boldsymbol{T_1^{-1}} \quad \textbf{[I.6]}$$

in which, $\boldsymbol{d} = \mu_0 h \gamma_A \gamma_X \langle r_{AX}^{-3} \rangle / (8\pi^2)$ and $\boldsymbol{c} = \omega_A \Delta\sigma/\sqrt{3}$, where μ_0 is the permeability of free space, h is the Planck's constant, γ_i is the gyromagnetic ratio of spin A and spin X, r_{AX} is the internuclear distance between A and X, ω the Larmor frequency of spin A or X, Δσ is the CSA of spin A, and $J(\omega)$ is the

spectral density function which measures the density of fluctuations at a frequency ω.

The dynamical information encoded by these parameters (T_1, T_2, and NOE) can be interpreted using the model-free approach developed by Lipari and Szabo (Lipari and Szabo 1982a; Lipari and Szabo 1982b) and extended by Clore and co-workers (Clore et al. 1990). As the name itself suggests, model-free does not assume any specific model for the molecular reorientation of spins. In this approach the spectral density function is modelled as:

$$J(\omega) = \frac{2}{5}\left[\frac{S^2\tau_{\mathrm{m}}}{1+(\omega\tau_{\mathrm{m}})^2} + \frac{(1-S_f^2\tau_f')}{1+\left(\omega\tau_f'\right)^2} + \frac{(S_f^2-S^2)\tau_s'}{1+(\omega\tau_s')^2}\right] \qquad \textbf{[I.7]}$$

in which τ_f' and τ_s' are the internal correlation times for fast and slow motions, respectively, τ_{m} is the isotropic rotational correlation time of the protein, $S^2 = S_f^2 S_s^2$ is the square of the generalized order parameter and characterizes the amplitudes of the internal motions ranging from zero for isotropic motion to one for completely restricted motion. A caveat is necessary as accurate interpretation of relaxation data with this expression requires that overall motion and internal motions are independent (*i.e.*, $\tau_f', \tau_s' \ll \tau_{\mathrm{m}}$).

Most commonly, experimental data is fitted to the parameters shown before using Monte Carlo simulations with initial estimates of τ_{m} obtained from T_1/T_2 ratios (Mandel et al. 1995; Palmer, III 2001). Presently, several software packages provide a complete model-free analysis of the relaxation data making a complex analysis accessible to non-experts. However, the versatility of these software packages is generally proportional to the difficulty in its use. A few examples include:

- *Modelfree* (Palmer's group), very versatile, allowing many options to be changed, although requiring many input files and limited output file formats (Mandel et al. 1995; Palmer et al. 1991);
- *DASHA*, contains an additional module for hydrodynamic calculations and allows PostScript plotting (Orekhov et al. 1995);
- *FAST-Modelfree*, (almost) fully automated analysis being one of the first to include a graphical user interface (Cole and Loria 2003);
- *Tensor2*, includes a very developed graphical user interface, providing clear visualization of the several steps of data analysis and easy plotting, however it is restricted to the analysis of ^{15}N relaxation data (Cordier et al. 1998; Dosset et al. 2000);
- *Relax*, is a rather recent software and very powerful due to python based scripts. It also presents the possibility of a new methodological analysis in the optimization of model-free parameters and rotational diffusion tensor. However, it requires good knowledge of python language - although a new version with an optional graphical user interface has just been released (d'Auvergne and Gooley 2008a; d'Auvergne and Gooley 2008b).

All these programs, have contributed to increase the accuracy in the determination of dynamic information from spin relaxation experiments and to make it accessible to more and more researchers as attested by the great increase in the number of studies published in the past decade.

Microsecond to millisecond chemical exchange

Chemical exchange occurs when nuclear spins are transferred between distinct local magnetic environments in a time dependent manner. The simplest and most common case is the two-site chemical exchange (R_{ex}) in which R_{ex} depends on the relative populations of the two states, the

respective chemical shift (ω), the rate of exchange (k_{ex}) and the effective field strength. The chemical shift difference (Δω) between the two states reflects the separation of the respective signals in the frequency spectrum expressed in Hz. The rate of exchange quantifies the number of exchange events per time unit hence is expressed in per second units (s^{-1}). The relation between Δω and k_{ex} allows the separation of chemical exchange in three time regimes: i) fast exchange, in which k_{ex} is much larger than the chemical shift difference between the two sites, ii) intermediate exchange, where k_{ex} and Δω are similar, and iii) slow exchange, characterized by a k_{ex} which is much slower than Δω.

The linewidths of the NMR resonances are sensitive to the dynamic properties of the molecule and relate to the spin-spin (T_2) relaxation rate. In the intermediate regime, chemical exchange broadens the linewidths of NMR signals increasing the measured T_2 relaxation rate. Line broadening results from the stochastic variations in the chemical shift induced by the exchange between the two sites. In this regime, the signal from one of the states (the higher energy low populated one) is usually weak and broadened beyond detection. However, the signal arising from the ground state spin population is still measurable with a line-broadening that reports on the exchange with the excited state. Importantly, line broadening due to exchange can be progressively suppressed by introduction of successive refocusing pulses in the NMR acquisition scheme (Figure I-5) (Palmer and Massi 2006; Baldwin and Kay 2009). Therefore, chemical exchange of a given nucleus can be measured by determining T_2 relaxation rates with decreasing contributions from chemical exchange effects (*i.e.*, increasing number of refocusing pulses). This approach is called relaxation dispersion and provides information on the thermodynamics and kinetics of the exchange reaction and the chemical shift values of the excited states – as a side note, it should be mentioned that

these chemical shift values allow structural information to be extracted for protein states which can represent as little as 0.5% of the whole population (Baldwin and Kay 2009).

Experimentally, relaxation dispersion experiments can be based on the Carr-Purcell-Meiboom-Gill (CPMG) pulse sequence for suppressing line broadening or on the rotating frame ($R_{1\rho}$) spin-lock. In general terms, the difference is that, in the first case the magnetization is kept aligned with the laboratory frame while in the second the spin-lock keeps the magnetization aligned with the rotating frame. However, in both cases exchange line-broadening is progressively suppressed thus increasing signal intensity as a function of the effective field strength. CPMG based experiments are instrumentally limited to maximum field strengths of about 1-3 kHz (the field strength being $1/4\tau$ with τ corresponding to the delay between refocusing pulses). Therefore, this approach is limited to chemical exchange measurements in regimes slower than about 200 microseconds. Higher frequency fields can be applied in $R_{1\rho}$ based relaxation dispersion experiments using spin lock pulses typically 4-5 fold stronger, allowing to extend chemical exchange studies to motions in the 100-20 microsecond time window - motions typically associated with loops and side-chains.

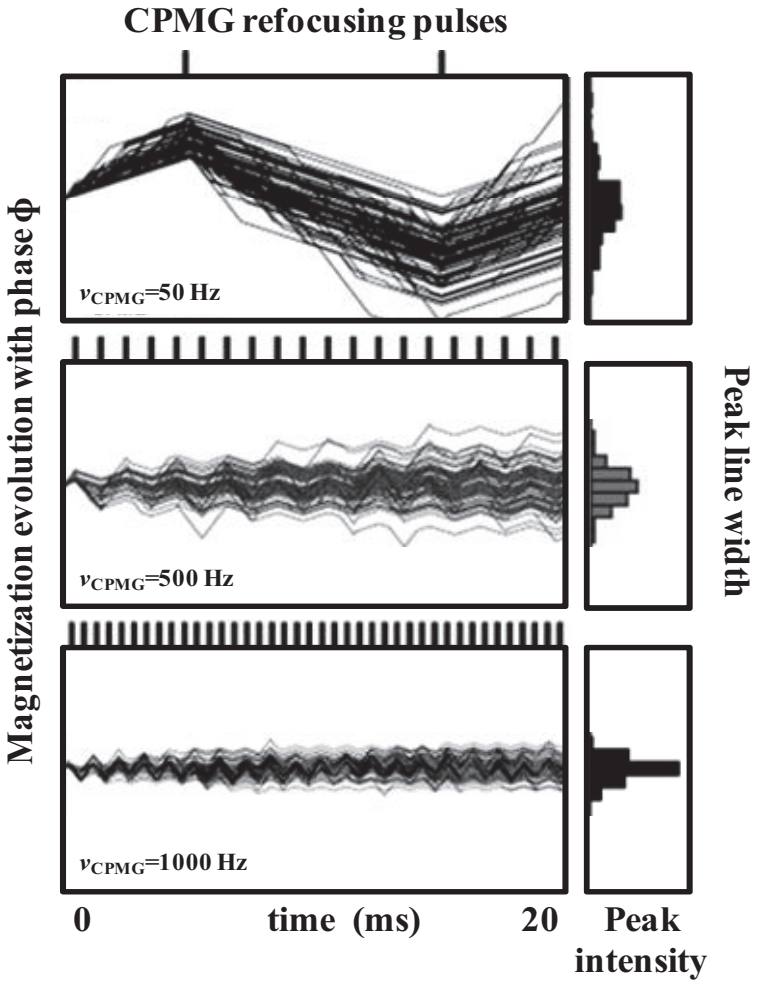


Figure I-5. Progressive suppression of exchange line broadening in a Carr-Purcell-Meiboom-Gill relaxation dispersion experiment. Lines represent magnetization evolution with phase (ϕ) as function of time for an ensemble of spins involved in stochastic chemical exchange. The slopes represent the slightly different chemical shift for each spin of the ensemble. Vertical bars at top of the plots represent refocusing pulses during the relaxation time which progressively suppress dephasing as the number of pulses increases (adapted from Mittermaier and Kay 2009).

Millisecond and beyond conformational motions

The rate of exchange of amide protons with water protons has been the most common approach to study conformational motions taking place in the millisecond to second and even slower time scales. The theoretical frame work of Hydrogen/Deuterium (H/D) exchange was established many years ago by the pioneering work of Linderstrøm-Lang and colleagues at the Carlsberg laboratory, namely John Schellman and Hvidt (Hvidt 1964; Hvidt and Nielsen 1966; Linderstrøm-Lang 1958; Schellman and Schellman 1997), and later by Englander's group (Bai et al. 1993; Molday et al. 1972). Hvidt was one of the main contributors to the field after Linderstrøm-Lang's death. She was responsible for the publication of the reaction depicted in I.8 although she acknowledges Linderstrøm-Lang as the author of the equation (Hvidt 1964). A very important advance in the understanding of the mechanism of hydrogen exchange was the finding that the exchange reaction was catalyzed by strong acids and bases (Berger et al. 1959).

A two state situation was assumed to describe the exchange reaction, with one of the states being hydrogen non-exchangeable (the "closed" state) and the other state being susceptible to hydrogen exchange (the "open" state) through a transient opening reaction of the protein structure:

$$\mathrm{NHclosed} \underset{k_{cl}}{\overset{k_{op}}{\rightleftharpoons}} \mathrm{NHopen} \xrightarrow{k_{ch}} \mathrm{Exchanged} \qquad \textbf{[I.8]}$$

in which k_{op} and k_{cl} represent the opening and closing rates of the transient structural fluctuation, and k_{ch} the exchange rate of a freely exposed amide proton. Under steady-state conditions (Linderstrøm-Lang 1958; Hvidt and

Nielsen 1966) and stable protein structure, that is assuming $k_{cl} >> k_{op}$, the exchange rate constant (k_{ex}) for reaction I.8 is given by equation I.9:

$$k_{\mathrm{ex}} = \frac{k_{\mathrm{op}} k_{\mathrm{ch}}}{k_{\mathrm{cl}} + k_{\mathrm{ch}}}$$

[I.9]

This equation can be further reduced to two limiting cases: the so called *EX1* and *EX2* conditions. In *EX1* conditions $k_{cl} << k_{ch}$, therefore the measured exchange is limited by the opening rate of the structural fluctuation, *i.e.*, $k_{\mathrm{ex}} = k_{\mathrm{op}}$. This regime is typical of high pH, high temperature, and low stability conditions and provides kinetic information about the transient opening reaction. At the *EX2* limit $k_{cl} >> k_{ch}$, and equation I.9 becomes:

$$k_{\mathrm{ex}} = \frac{k_{\mathrm{op}}}{k_{\mathrm{cl}}} k_{\mathrm{ch}} = K_{\mathrm{op}} k_{\mathrm{ch}}$$

[I.10]

The inverse of K_{op} is called the protection factor and expresses the degree of structural protection against exchange. Even in the absence of folded protein structure, side-chains affect the exchange rate of freely exposed amide protons (k_{ch}). Thus, k_{ch} depends on the pH (by acid/base catalysis) and temperature but also on the local amino acid sequence. The values of k_{ch} have been determined to high precision using peptide models for the different side-chains (Bai et al. 1995), allowing the determination of the opening constant, K_{op}. The free energy value (ΔG_{HX}) for the structural opening reaction I.8, can then be calculated as:

$$\Delta G_{\mathrm{HX}} = -RT\ ln\ K_{\mathrm{op}} = -RT \ln\left(k_{ex}/k_{ch}\right)$$

[I.11]

As exchange is acid/base catalyzed, the *EX2* limit condition can be confirmed if a ten-fold variation in the exchange rate is observed when the pH is changed by one unit. It is important to note that in this regime, changes in the measured exchange rates reflect a shift in the equilibrium of the folded and unfolded protein populations. However, proton/deuterium exchange depends on motions that expose the hydrogens to the solvent. Therefore, a reduction of the exchange rates is interpreted as a decrease in the amplitudes and/or number of motions of the native state.

In a typical H/D exchange experiment, the reaction is initiated by dissolving lyophilized protein in a 99% D_2O solution. As amide protons from the molecule exchange with the excess of deuterium atoms from the solvent, according to equation I.8, the NMR signal of each amide proton decays over time with a rate that can be determined by acquiring consecutive NMR spectra, (usually ^{1}H-^{15}N-HSQC) and measuring peak intensities. When the opening motions in reaction I.8 become much faster than 10^{-3} to 10^{-6} s^{-1}, conventional H/D exchange experiments are no longer suitable since the experimental dead time becomes greater than the total H/D exchange time. Alternatively, magnetization transfer techniques coupled with fast HSQC acquisition schemes are most appropriate to measure exchange rates in the range of 2-50 s^{-1} (Hvidt and Nielsen 1966; Hwang et al. 1997; Mori et al. 1995; Mori et al. 1996).

Hydrogen exchange measurements can also report on the amplitude of the motion involved in the mechanism of exchange, *i.e.*, amide exchange may result from local protein motions or more global motions. To distinguish between these different scale motions, the dependence of exchange rates on low denaturant concentrations has to be determined (Mayo and Baldwin 1993). The slope of such dependence is called the *m*-value and was proposed to correlate with the increase in surface exposure resulting from the opening

motion of the protein (Tanford 1970; Schellman 1987a) - although significant debate surrounds this matter (Wooll et al. 2000). Amide exchange via local motions occurs with little new surface area of the protein exposed to the solvent, making them relatively insensitive to the denaturant concentration (*m*-value close to zero); on the other hand, exchange via global motions occurs with a significant increase in the protein surface area therefore exposing many additional denaturant binding sites. As a consequence global motions become rapidly favoured with the increasing concentration of the denaturant. This approach is particularly interesting as it allows probing intermediate states that are only transiently populated and informs about the folding/unfolding pathway of a protein (Krishna et al. 2004; Bai et al. 1995). Englander and colleagues have successfully used this approach to identify partly unfolded intermediates and the respective folding/unfolding sequence of the cytochrome *c* and staphylococcal nuclease proteins (Bai et al. 1995; Bedard et al. 2008b).

In conclusion, hydrogen exchange allows obtaining information about slow motions of proteins but it also provides information about the extent of these motions within the protein structure.

SOLUTE INDUCED EFFECTS ON PROTEIN DYNAMICS

The study of the effect of stabilizing solutes on protein dynamics has been less explored, especially in comparison to the number of studies on the thermodynamic analysis. Also, the tested compounds have been mainly neutral or zwitterionic organic solutes like sucrose, glycerol, glycine, and trimethylamine *N*-oxide (TMAO). Studies with charged compounds have focused mostly on the effects of inorganic ions with little information gathered about the effect of charged organic solutes on protein dynamics (Lamosa et

al. 2003). Concerning the time scales, it has been most common to monitor the effect of solutes on the slowest and on the fastest motion time scales using hydrogen/deuterium exchange and spin relaxation measurements, respectively.

Despite the limited number of observations, solutes that stabilize proteins have been generally found to constrain mainly the slow, global motions of proteins while destabilizers (*e.g.,* urea) have been found to promote them. For example, sucrose (1 M), glycerol (45 and 75%), and glycine (2 M) were found to induce a significant decrease on the hydrogen/deuterium exchange rates of the most slowly exchanging amide protons of ribonuclease A, myoglobin, and cytochrome *c* /chymotrypsin inhibitor 2, respectively, while leaving faster exchange rates virtually unaffected (Calhoun and Englander 1985; Foord and Leatherbarrow 1998; Wang et al. 1995). This preferential effect of the solutes on the amide protons with the slowest exchange rates was interpreted as a decrease in the amplitudes and/or number of fluctuations that involve large increases in the surface exposure of proteins, *i.e.*, global or nearly global motions. Since stabilizing solutes establish unfavourable interactions with the protein backbone, global motions that result in greater exposure of the backbone to the solute will be more disfavoured than local motions with little contribution to the surface exposure increase. In the specific case of TMAO, a decrease in hydrogen exchange rates in the presence of this solute was also observed for the amide protons of cold shock protein A and ribonuclease A, however the most significant effects were not restricted to the slower exchanging protons (Jaravine et al. 2000; Qu and Bolen 2003). For the cold shock protein A, a rather unstable protein (3.9 kcal/mol at 5°C), exchange rates of many weakly protected amide protons could only be measured in the presence of 0.2 M TMAO, indicating that significant restriction occurs also for amide protons with faster exchange rates (Jaravine et al. 2000). On the other

hand, the faster exchange rates of the ribonuclease A amide protons were almost unaltered in the presence of 1.0 M TMAO (Qu and Bolen 2003). Concerning the effects of inorganic salts on the amide proton exchange rates, it was found that at moderate anion concentrations (0.2 – 1 M) exchange rates are modulated according to the Hofmeister series, *i.e.*, anions that are most excluded from the protein surface increase the water surface tension and therefore induce a restriction on protein motions (Tadeo et al. 2007). Another study monitored the amide proton exchange rates of rubredoxin, a small protein from *Desulfovibrio gigas*, in the absence and presence of diglycerol phosphate (0.1 M), but no clear match was found between the magnitude of exchange rate reduction and the slowest exchanging protons (Lamosa et al. 2003). In fact, the effect was quite variable with some fast exchanging amide protons being more affected than slower exchanging ones. Importantly, even at such low concentrations, the solute is found to induce a fivefold decrease in the exchange rates of many amide protons. Therefore, it appears that the slowest exchanging amide protons are always affected by stabilizing solutes suggesting a more compact protein structure. On the other hand, the effect on the amide protons with faster exchange rates seems unclear.

With respect to solute induced changes on the faster motions of the protein, that is, from the microsecond to the picoseconds time scales, the few studies reported until now suggest a tendency for decreased backbone mobility in the presence of stabilizing solutes and the opposite effect in the presence of destabilizers. For example, the increase in the fast backbone motions of ribonuclease A brought about by the presence of 0.7 M guanidine hydrochloride is almost completely reverted upon the addition of 0.35 M TMAO (Doan-Nguyen and Loria 2007). The authors suggest that the most pronounced effects of guanidine hydrochloride occur at positions of the

protein believed to be the first to unfold. Another study reports a slight trend for increased rigidity of the rubredoxin fast motions in the presence of diglycerol phosphate. However, no correlation was found between the magnitude of the effect and the structural properties of the protein. Spin labelled proteins containing several nitroxide side-chains were used to probe the effect of sucrose (30% w/w) on exchange motions with life-times slower than 100 nanoseconds (Lopez et al. 2009). The results suggested that the most mobile regions of the protein were more restricted by the presence of sucrose. Whether the proposed correlations hold for other protein models and solutes remains naturally dependent on further studies.

In summary, the present studies suggest that protein motions at more than one time scale can be restricted by the presence of stabilizing solutes. Furthermore, there is some evidence supporting the idea that the rigidification induced by the solutes goes together with protein stabilization. However, no clear conclusions have emerged about the relevance of each time scale to protein stabilization or about the magnitude of the effect in different regions of the protein. A multiple time scale approach in a unique protein model would be very useful in assessing the importance of each time scale to the protein stabilization observed in the presence of solutes. Furthermore, studying protein motions as a function of solute concentration provides a more solid base to search for a correlation between changes in protein dynamics and changes in protein stability induced by solutes.

Chapter II

Structural determinants of protein stabilization by solutes: the importance of the hairpin loop in rubredoxins

The results of this chapter were published in:
Pais T.M., Lamosa P.L., dos Santos W., LeGall J, Turner D.L., Santos H. (2005) Structural determinants of protein stabilization by solutes: the importance of the hairpin loop in rubredoxins. *FEBS J.* 272:999-1011.

CONTENTS

SUMMARY

Despite their high sequence homology, rubredoxins from *Desulfovibrio gigas* and *Desulfovibrio desulfuricans* are stabilized to very different extents by compatible solutes such as diglycerol phosphate, the major osmolyte in the hyperthermophilic archaeon *Archaeoglobus fulgidus* (Lamosa et al. 2000). The principal structural difference between these two proteins is the absence of the hairpin loop in the rubredoxin from *D. desulfuricans.* Therefore, mutants of *D. gigas* rubredoxin bearing deletions in the loop region were constructed to investigate the importance of this structural feature on protein intrinsic stability, as well as on its capacity to undergo stabilization by compatible solutes. The three-dimensional structure of the mutant bearing the largest

deletion, Δ17|29, was determined by ^{1}H-NMR, demonstrating that, despite the drastic deletion, the main structural features were preserved. The dependence of the NH chemical shifts on temperature and solute concentration (diglycerol phosphate or mannosylglycerate) provide evidence of subtle conformational changes induced by the solute. The kinetic stability (as assessed from the absorption decay at 494 nm) of six mutant rubredoxins was determined at 90°C and the stabilizing effect exerted by both solutes was assessed. The extent of protection conferred by each solute was highly dependent on the specific mutant examined: while the half-life for iron release in the wild-type *D. gigas* rubredoxin increased threefold in the presence of 0.1 M diglycerol phosphate, mutant Δ23|29 was destabilized. This study provides evidence for solute-induced compaction of the protein structure and occurrence of weak, specific interactions with the protein surface. The relevance of these findings to our understanding of the molecular basis for protein stabilization is discussed.

INTRODUCTION

In spite of the extensive accumulation of data on protein structure, the molecular determinants of protein thermal stability remain elusive. Also, the beneficial stabilizing effects exerted by various compatible solutes have been known for a long time, yet the mechanisms responsible for this stabilization are a matter of intense discussion (Batchelor et al. 2004; Bolen 2001; Qu and Bolen 2003; Timasheff 2002). One of the reasons for this apparent lack of success is that many different factors, both intrinsic and extrinsic, seem to contribute to the thermostability of any given protein (Petsko 2001). Protein stability appears as the result of a delicate balance of stabilizing and destabilizing interactions, with the thermodynamic stability of the native state

emerging as a small difference of large numbers(Jaenicke and Bohm 1998). Similarly, the stabilizing effect conferred by compatible solutes will be the result of a plethora of direct and/or indirect, weak interactions between the solute (or the changes that the solute causes in the solvent properties) and the several chemical groups present on the protein surface, rendering the magnitude of this effect subtly dependent on the particular solute/protein pair examined and, therefore, extremely difficult to predict.

One of the strategies used to explore this maze of interactions and try to rationalize them is to investigate series of homologous proteins in order to unravel the structural determinants of protein stabilization by compatible solutes. In a previous study we compared the action of a compatible solute, diglycerol phosphate (DGP), on the stability of rubredoxins from three bacterial sources (Lamosa et al. 2000). These small metalloproteins display a wide variation in thermal stability, despite having a considerable degree of sequence and structural similarity. Typically, rubredoxins are composed of about 52–54 residues and include a three-stranded β-sheet, a metal centre comprising one iron atom tetrahedrally coordinated by four cysteine sulfur atoms, and a small hydrophobic core, which is shielded from solvent access by a hairpin loop (Sieker et al. 1994). Despite the structural similarity between rubredoxins, the degree of stabilization conferred by diglycerol phosphate was diverse. Although having almost no effect on the thermal stability of the rubredoxin (Rd) from *Desulfovibrio desulfuricans* (RdDd), diglycerol phosphate was able to triple the half-life for thermal denaturation of the other two rubredoxins examined. RdDd is the least heat-stable of the several rubredoxins investigated, and is the only one not stabilized by diglycerol phosphate. Conversely, the Rd from *D. gigas* is the most stable and strongly stabilized by this solute. The main structural difference between RdDd and other rubredoxins is the lack of seven amino acids in the hairpin loop.

In order to investigate why this structural feature (the presence of the loop region) seemed to have such a profound effect on the stability and stabilization of rubredoxins, we constructed a series of mutants of the Rd from *D. gigas* with different extents of deletion in the original hairpin loop. The determination of the NMR solution structure was deemed important, first, to ensure that the deletion had not substantially altered the protein structure (except in the loop region); and second, to provide the structural detail needed to elucidate the molecular basis of protein stabilization by solutes. Three point mutants were also studied to assess the importance of total surface charge or changes in the most exposed hydrophobic residue.

Diglycerol phosphate and mannosylglycerate (MG), two negatively charged compatible solutes that we isolated from hyperthermophiles, were used in this study. The effect of these solutes on the thermal stability of six mutants was investigated. Moreover, as chemical shifts are good indicators of changes in protein structure or dynamics, the changes of the proton chemical shifts with temperature and solute concentration were analysed to extract information on protein/solute interactions.

MATERIALS AND METHODS

Mutagenesis of Rd and protein production in *Escherichia coli*

Plasmid pRPPL1 (Lamosa et al. 2000), harbouring the gene encoding RdDg under the control of a heat sensitive promoter, was constructed from pCYTEXP1 (Belev et al. 1991) and used as a template for all mutations. Mutants Δ17|29, Δ23|29 and Δ17|26 were obtained using the ExSite™ PCR-Based Site-Directed Mutagenesis Kit according to the procedures outlined in the respective kit instructions (Stratagene, La Jolla, CA, USA). Mutants D2K and K17E were constructed to evaluate the effect of total surface charge on

the stabilizing efficacy of the solutes. A third mutant (V8N) was designed to assess the effect of changing the most exposed hydrophobic residue of the native structure**.** These mutants were constructed using the QuikChange™ Site-Directed Mutagenesis Kit (Stratagene). The sequence of the coding unit of each mutant was confirmed by restriction analysis of DNA isolated from positive recombinant clones. The resulting plasmids, encoding mutant proteins D2K, K17E and V8N, were cloned in *Escherichia coli* strain XL1BLUE. This system led to very low yields for the expression of mutants Δ23|29, Δ17|26 and Δ17|29, hence the respective coding units were transferred to an isopropyl thio-β-d-galactoside-inducible plasmid, pt7-7 (Lamosa et al. 2000).

Mutant proteins were purified using three chromatographic steps as previously described (Lamosa et al. 2000). After the last column in the purification procedure (Resource Q; Pharmacia, Uppsala, Sweden), iron (red) and zinc (colourless) forms of rubredoxin mutants were separated and judged pure by silver-staining native PAGE electrophoresis.

Thermal stability assays

The kinetics for the disruption of the iron Rd structure at 90 °C was monitored by UV–visible absorption spectroscopy in a Shimadzu UV-1601 spectrophotometer equipped with a thermostated cell. A rubber septum was adapted to a quartz cell to allow measurements under anaerobic conditions (subjected to three vacuum-argon cycles of 15 minutes). The assay solution consisted of 50 mM Tris/HCl buffer, pH 7.6, and the desired concentration of a given solute. The temperature of the solution in the spectrophotometric cell was measured with a thermocouple. Once thermal equilibrium was reached, the desired amount of protein solution (c.a. 20 μM final concentration) was rapidly added and spectral scanning started. Spectra were recorded for each time point and baseline corrected. The values of absorbance measured at 494

nm (A_{494}) as a function of time were fitted to a single exponential decay. The effect of 0.1 M diglycerol phosphate and 0.2 M mannosylglycerate on the degree of stabilization of the native RdDg was studied (above these concentrations there was no significant increase in the magnitude of stabilization).

NMR structure calculation and analysis

^{1}H-NMR spectra of zinc rubredoxin Δ17|29 were acquired on a Bruker DRX500 (Bruker, Rheinstetten, Germany) spectrometer equipped with a 5 mm probe head with internal B_0 gradient coils. Protein NMR samples (4 mM) were prepared in 10% $^{2}H_2O$ and the pH adjusted to 7.6. Assignment of the proton signals was performed in spectra acquired at 303 K, but additional spectra were obtained at 313 K to help resolve peak overlap, especially in the aromatic region. NOESY spectra (mixing times of 35, 70, 80 and 100 ms), total correlation spectra (TOCSY) (Briand and Ernst 1991) using the clean total correlation spectroscopy pulse sequence with spin lock times of 70 and 100 ms, and COSY spectra were recorded. Raw data was processed using standard XWIN-NMR software (Bruker). Polynomial baseline corrections were applied in both dimensions of each spectrum. The software XEASY (v. 1.3.10) was used for assignment and integration of NOESY cross-peak volumes. NOEs were measured at 303 K in the 80 ms NOESY spectrum.

The NOESY cross-peak volumes were used to calculate upper (upl) and lower (lol) limit volumes using the program INDYANA (Turner et al. 1999). Nonstereospecifically assigned protons, degenerate protons, overlapping peaks, and flexible proline rings were treated as previously described (Turner et al. 1999; Brennan et al. 2000). Stereospecific assignments were determined with the help of the program GLOMSA (Brennan et al. 2000; Guntert et al. 1991). Upper and lower distance limits for each pair of sulfur atoms involved

in zinc coordination (C6, C9, C26 and C29) were fixed between 3.9 and 3.5 Å (Sieker et al. 1994; Lamosa et al. 2001; Dauter et al. 1996; Brennan et al. 2000). This range of distances allows a significant distortion from tetrahedral geometry and does not specify the chirality of the centre. The experimental distance restraints were then used as input to generate protein conformers using the program for restrained dynamics and simulated annealing, DYANA v. 1.4 with modifications (INDYANA) to optimize scaling factors for calibrating NOE intensities (Brennan et al. 2000). In the final refinement stages, each batch of structures was checked for the existence of short distances (<2.5 Å) between protons for which no NOE had yet been measured. If an expected NOE was not visible in a clear region of the spectra, the volume at that frequency was measured and used to provide a lower limit distance in subsequent structure calculations, thus reducing the possibility of incorrect short inter-proton distances. A complete relaxation matrix analysis was applied to an ensemble of 10 structures to estimate the error that might be introduced via spin-diffusion; this value was used to loosen all distance constraints in subsequent calculations (Brennan et al. 2000).

The program MOLMOL v. 2.6 (Koradi et al. 1996), was used for superimposition and visual inspection of the final family of structures. The NMR structures were analysed with respect to experimental constraints using standard procedures of the DYANA program. The quality of structures with respect to dihedral angles was evaluated using the program PROCHECK-NMR v. 3.4.4 (Laskowski et al. 1996) and Ramachandran plots were generated. The optimal hydrogen bond network was calculated for each structure using program WHAT IF v. 5.0 (Rodriguez et al. 1998). The atomic coordinates and constraint files were deposited in the Protein Data Bank at the Research Collaboratory for Structural Informatics-Rutgers with the accession code 1SPW.

Proton chemical shift variation

Samples of mutant Zn rubredoxin Δ17|29 with a concentration of 0.4 mM and pH 7.6 were prepared in 10% ^2H_2O. Chemical shifts were followed in TOCSY spectra (Briand and Ernst 1991) obtained with water presaturation and phase-sensitive mode using TPPI. Spectra were acquired in the presence of different concentrations of diglycerol phosphate, KCl, trehalose and mannosylglycerate. Temperature dependence of proton chemical shifts was evaluated between 273 and 303 K at 2.5-degree intervals in the absence of solutes and in the presence of 200 mM diglycerol phosphate, 200 mM KCl or 400 mM mannosylglycerate. This dependence was expressed by the respective temperature coefficient (10^{-3} ppm·K^{-1}).

Proton chemical shifts are referenced to internal 3-(trimethylsilyl)propanesulfonic acid (sodium salt), and analysed using XEASY (v. 1.3.10) software. Secondary shifts were obtained by subtraction of random coil chemical shifts (Wüthrich, 1986) from the experimental ones.

RESULTS

Thermal stability of rubredoxins

The mutant iron rubredoxins show the same characteristic bands of the UV-visible absorption spectrum as the native protein with maxima centred at 380, 494 and 570 nm. These bands are bleached due to the disruption of the iron centre when the protein undergoes denaturation. Monitoring the loss of the metal centre through the decrease in absorbance at 494 nm provides an expeditious way to evaluate the kinetic stability of rubredoxins (Cavagnero et al. 1998b; Cavagnero et al. 1998a; Eidsness et al. 1997). The half-life ($t_{1/2}$) for iron release of the native and mutant rubredoxins was measured at 90°C.

All rubredoxins examined exhibited mono-exponential behaviour in regard to the decay of absorbance at 494 nm (data not shown). Complete bleaching of spectral features at 380 and 494 nm occurred without formation of detectable precipitates, either from protein precipitation or insoluble ferric oxides. The spectral features did not recover on cooling, which indicates that protein denaturation under these conditions is an irreversible process, in agreement with previous studies regarding thermal denaturation of rubredoxins (Cavagnero et al. 1998b; Eidsness et al. 1997; Lamosa et al. 2000).

Recombinant RdDg presented a half-life for disruption of the iron centre of 96 min with all mutations resulting in a decrease of this parameter. The mutants bearing deletions in the loop region showed a dramatic decrease (between 69 and 89%) in their half-lives relative to the native form (Table II-1 and Figure II-1).

Table II-1. Effect of solute addition on the half-life values (in minutes) for the thermal denaturation of native rubredoxins and mutants at 90°C

Protein	No solutes	DGP 0.1 M	MG 0.2 M
RdDg[a]	96.2 ±9.4	295.0 ±7.1	129.6 ±5.2
Δ17|29[b]	10.5 ±1.5	16.0 ±5.0	28.3 ±0.9
Δ17|26	14.1 ±1.4	15.9 ±2.1	25.2 ±5.8
Δ23|29	29.7 ±3.8	15.1 ±2.5	45.0 ±2.1
D2K	77.6 ±5.5	150.7 ±4.6	
K17E	55.5 ±4.1	98.7 ±6.7	
V8N	33.5 ±2.1	36.5 ±2.1	
RdDd[c]	30 ±4.0	35.7 ±4.0	

[a] The half-life in the presence of 0.2 M KCl is 104±13 min. [b] The half-life values in the presence of 0.2 M KCl and 0.4 M trehalose are 11.2±2.1 min and 19±1.7 min, respectively.[c] Values from Lamosa et al. 2000.

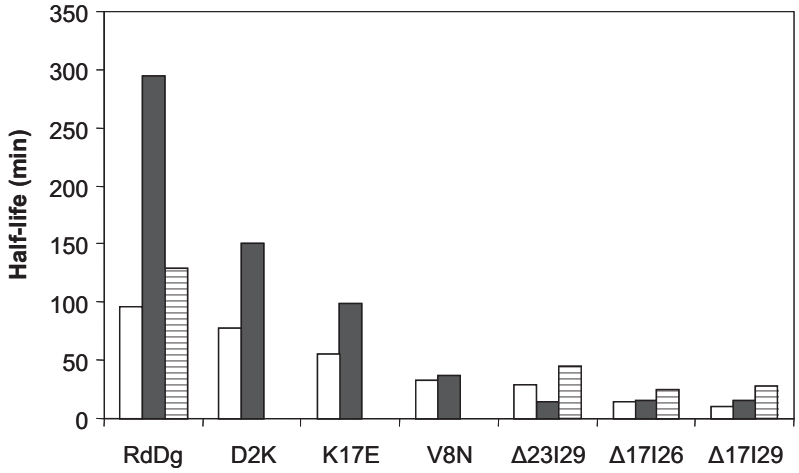


Figure II-1. Effect of DGP and MG on the thermal stability of *D. gigas* rubredoxin and several mutants. The half-life values for the thermal denaturation of proteins in the absence of solutes (empty bars), with 0.1 M DGP (solid bars) or with 0.2 MG (striped bars) are depicted.

Interestingly, mutant Δ23|29 had a half-life comparable with that of the Rd from *D. desulfuricans*, but the two other mutants lost iron at an even higher rate. The larger the deletion, the shorter the half-life became, with mutant Δ17|29 showing the lowest value for this parameter. In general, single mutations had a smaller effect on the rate of iron loss, except for V8N, which showed a rate comparable with that of mutant Δ23|29.

The effect induced by diglycerol phosphate (DGP) in native RdDg was impressive with at least a threefold increase of the half-life (Lamosa et al. 2000). However, the effect observed for the mutant rubredoxins was lower. Mutants D2K, K17E and Δ17|29 showed a clear increase in their half-life values (between 52 and 94%) but only a minor change was observed with mutants Δ17|26 and V8N (Figure II-1, Table II-1). Most surprisingly, the half-life of mutant Δ23|29 was reduced in the presence of diglycerol phosphate. It is also interesting to note that, for the point mutants, the added stabilization follows the intrinsic stability, with the larger increases occurring in the proteins

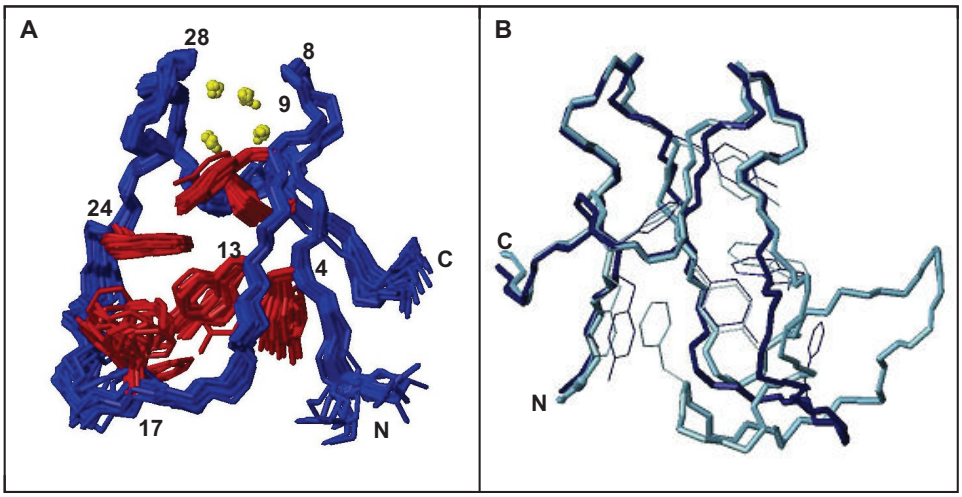
with higher intrinsic stability. This trend, however, was not observed in the case of loop deletions. The presence of mannosylglycerate caused a consistent retardation on the rates of all rubredoxins examined; in the case of RdDg the effect was much smaller than that of diglycerol phosphate, which contrasts with the greater effect observed for the deletion-mutants, including Δ23|29 (destabilized by DGP). Because K^+ was the counter-ion for the negative charge of diglycerol phosphate and mannosylglycerate, the effect of KCl on the rate of iron release was also determined. We found that KCl had no significant effect on the half-life of the proteins examined (Table II-1).

Structure determination of mutant Δ17|29 by NMR

Proton signal assignment was performed using the classical approach described by Wüthrich (1986). Analysis of TOCSY and COSY spectra allowed the identification of the spin systems. Sequence-specific assignment was achieved using NOESY spectra and identifying connectivities between NH protons and between the NH and H protons of adjacent spin systems. The spin-systems for Met1 and Asp19 could not be identified, probably because mobility of the N-terminus and the loop region leads to weak signals. Spin diffusion was taken into account and a value of 6.2% was used to loosen all NOESY-derived constraints. Stereospecific assignments were obtained using preliminary calculated structures with the aid of program GLOMSA; of these, 16 were derived from stereopairs with nondegenerate chemical shifts and 50 NOESY cross-peaks could be pseudo-stereospecifically assigned to one or the other side of the fast-flipping aromatic side-chain rings.

The program INDYANA was used to generate 500 conformers from which the 20 structures with the lowest target functions were selected. A schematic representation of the 20 superimposed structures showing the backbone, aromatic side-chains and cysteine sulfur atoms, is presented in Figure II-2A

and a statistical analysis is given in Table II-2. The metal centre conserves both the geometry and the chirality of the native protein and is well defined, with the heavy atoms of the four coordinating cysteines (residues 6, 9, 26 and 29) having an RMSD <0.55 Å. Analysis of the secondary structure with MOLMOL v. 2.6 (Koradi et al. 1996) and PROCHECK-NMR showed the presence of a three-stranded β-sheet similar to that of the native protein (Figure II-2B). The Ramachandran plot shows that most of the residues (94.7%) fall in the most favoured and additionally allowed regions; however, 5.2% appear in the generously allowed and one residue (Asp19) appears in the disallowed region in one of the 20 structures. This residue is located in the residual loop of the mutant and, if only well-defined regions are considered (Table II-2), no residue appears in the disallowed region. The deviation is probably a consequence of the large deletion (25% of the residues were deleted) straining the backbone to accommodate the conserved structural features.



C

	5	10	15	20	25	30	35	40	45	50
RdDg	M D I Y V C T V C G Y E Y D P A K G D P D S G I K P G T K F E D L P D D W A C P V C G A S K D A F E K Q									
RdDd	M Q K Y V C N V C G Y E Y D P A E H D - - - - - - - N V P F D Q L P S S W C C P V C G V S K D Q F S P A									
Δ17|29	M D I Y V C T V C G Y E Y D P A - - - - - - - - - - - - - F E D L P D D W A C P V C G A S K D A F E K Q									
	5	10	15			17	22	27	32	37

Figure II-2. NMR structure of mutant Δ17|29 and comparison with the structure of native D. gigas rubredoxin. A) 20 best NMR structures calculated for the mutant showing the backbone (blue), aromatic heavy side-chains (red) and cysteine sulfur atoms (yellow). B) Superimposition of the native (light blue) and mutant (dark blue) backbone and aromatic side-chains using all residues except those in the loop region. C) Sequence alignment of the mutant Δ17|29 and the native rubredoxins from *D. gigas* and *D. desulfuricans*; the upper numbering refers to the RdDg amino acid sequence while the lower refers to the mutant.

Table II-2. Restraint violations and quality analysis for the rubredoxin Δ17|29 mutant structure

DYANA target function		**Ramachandran plot (%)[a]**	
Average total (Å)	0.21 ±0.021	Most favoured	58.3(54.5)
Function Range	0.16 – 0.23	Additionally allowed	36.5(40.2)
Violated Constraints		Generously allowed	5.2(5.2)
Consistent violations (>0.2 Å)	0	Disallowed	0(0.2)
van der Waals (>0.2 Å)	0		
Precision (Å)		**Nonredundant distance restraints (lower limits)**	
Mean global backbone RMSD	0.98 ± 0.21	Intraresidual	109
Mean global heavy atom RMSD	1.72 ± 0.26	Sequential (|i-j|=1)	102
		Medium range (2=|i-j|<5)	92
		Long range (|i-j|>5)	138
		Total redundant and nonredundant	734

[a] Residues with S (Φ) and S (Ψ)< 0.8 were not included for the Ramachandran plot calculation; the values obtained using all residues are shown in brackets

Overall, the structure of the mutant retains the main features of the native structure with the obvious exception of the loop region. The RMSD between the backbones of the mean structures for the native and mutant rubredoxins is 2.24 Å. However, if residues 16–22 (sequence numbering of the mutant), which make up the shortened loop region in the mutant, are excluded, the deviation decreases to 0.82 Å, showing that this large deletion left the remaining structure virtually unaltered (Figure II-2B). The optimal hydrogen bond network was calculated for each of the 20 structures and it is also similar to that displayed by the native protein (Lamosa et al. 2001). However, the average exposure to water increased, especially in the segment 16–23, with

values over 40% observed for some of these residues (Figure II-3). In particular, the exposure of the residues that comprise the lower part (relative to the orientation depicted in Figure II-2) of the hydrophobic core of the native protein, namely, Y4, Y13, F17, L20 and W24 (numbering according to the mutant) increased substantially.

The structure of mutant Δ17|29 shows considerable similarities with RdDd, a protein naturally truncated in the loop region. In fact, excluding residues 16-22 (sequence numbering of the mutant), the RMSD between the backbones of the X-ray model of RdDd and the mutant rubredoxin is only 1.22 Å. However, if the residues corresponding to the residual loop region of mutant Δ17|29 are included, the RMSD between its backbone and that of RdDd increases to 2.64 Å. The most striking difference between mutant Δ17|29 and RdDd is the absence of a histidine residue in the mutant protein and the 6.2 Å shift of Phe17 (30 in the RdDd sequence).

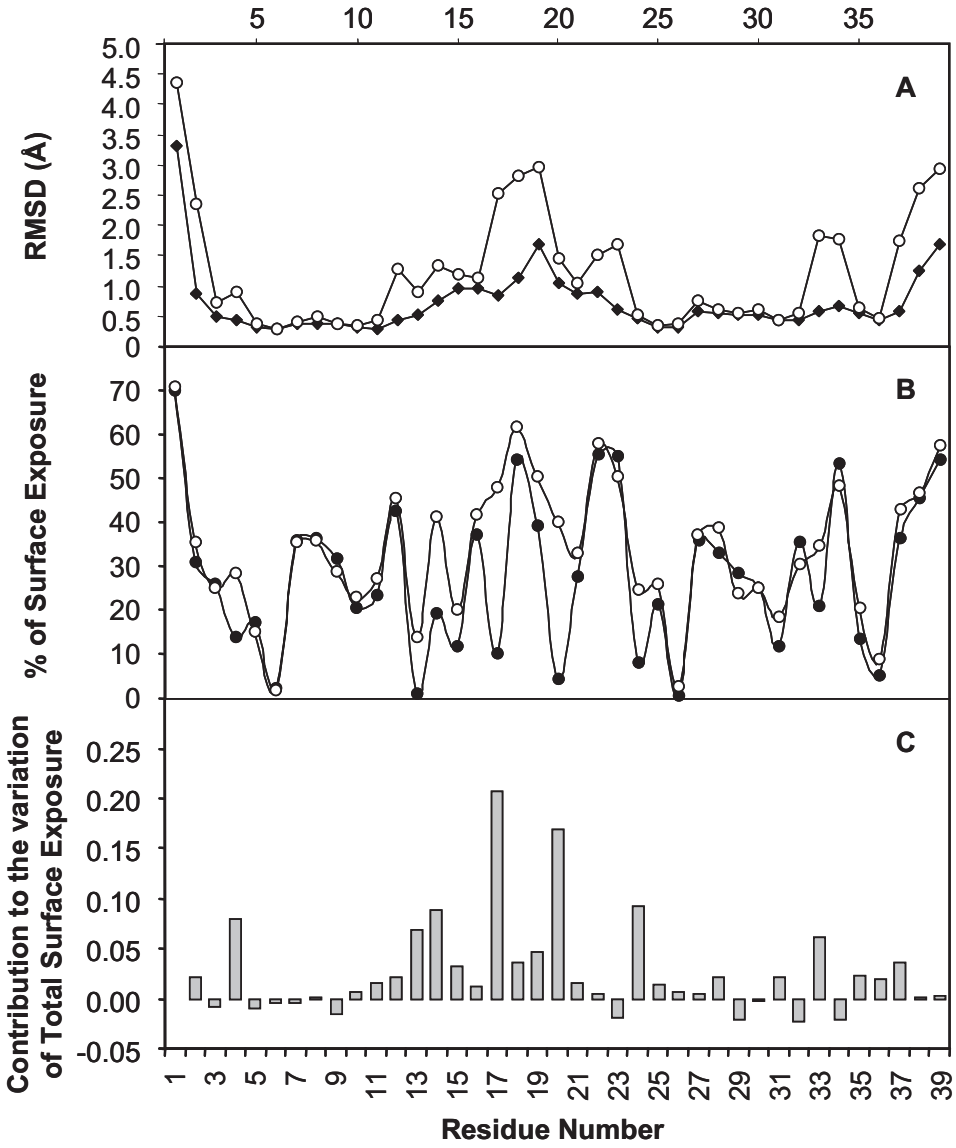


Figure II-3. Average RMSD values for each residue and respective average surface exposure. (A) RMSD values for the backbone (full diamonds) and heavy atoms (open circles). (B) Percentage of surface exposure per residue of mutant Δ17|29 (open circles) and wild-type *D. gigas* rubredoxin (RdDg) (full circles). (C) contribution to the variation of the total surface exposure of the mutant protein with respect to the wild-type RdDg.

Dependence of chemical shifts on solute concentration

Chemical shifts are sensitive probes of protein conformation. Thus, in an effort to explore possible structural alterations that solutes might induce in the protein, or preferential interactions with specific protein loci, the chemical shifts of all assigned protons in Δ17|29 zinc rubredoxin were measured in the presence of different solute concentrations. Variation of NH chemical shifts along the protein backbone demonstrated an intriguing pattern common to diglycerol phosphate and mannosylglycerate (Figure II-4) with the major shift variations occurring in the truncated loop region. This led to two hypotheses: either the action of both solutes upon the structure was very similar, or the observed shifts were a consequence of increasing ionic strength.

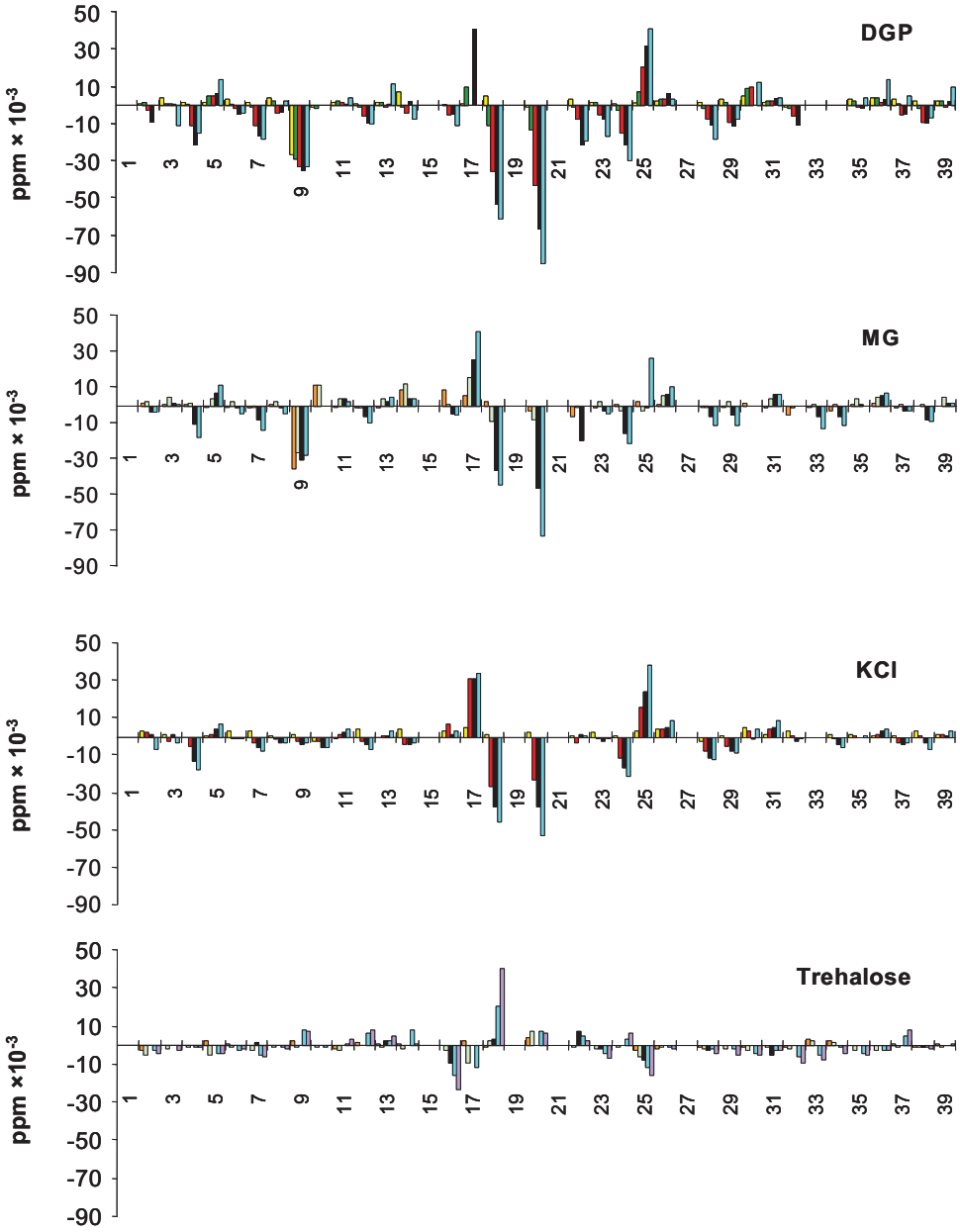


Figure II-4. Dependence of the NH chemical shifts for mutant Δ17|29 with solute concentration. Bars are arranged from left to right at increasing solute concentrations: 1 mM (), 2 mM (), 10 mM (), 20 mM (), 50 mM (), 100 mM (), 200 mM () and 400 mM ().

To distinguish between these two hypotheses, KCl (a charged solute with no significant effect on the half-life values) and trehalose (an uncharged solute that retards iron loss) were also used to measure chemical shift variations. KCl, used to control for ionic strength effects, presented a pattern that is very similar to that observed with the other charged solutes, while the effect of trehalose, although concentrated in the same region, induced much smaller shifts which tended to be of the opposite sign (Figure II-4). However, after discounting for the ionic strength effect, the shifts observed in the presence of diglycerol phosphate and mannosylglycerate (Figure II-5) become comparable in size with those displayed with trehalose.

Significant shifts were also observed for other types of protons upon solute addition. However, these were not monotonic with solute concentration and showed no obvious pattern. Correlation between experimental chemical shifts and several parameters, such as solvent exposure, RMSD, secondary shift and temperature coefficients were also analysed but no obvious good correlation was found (not shown).

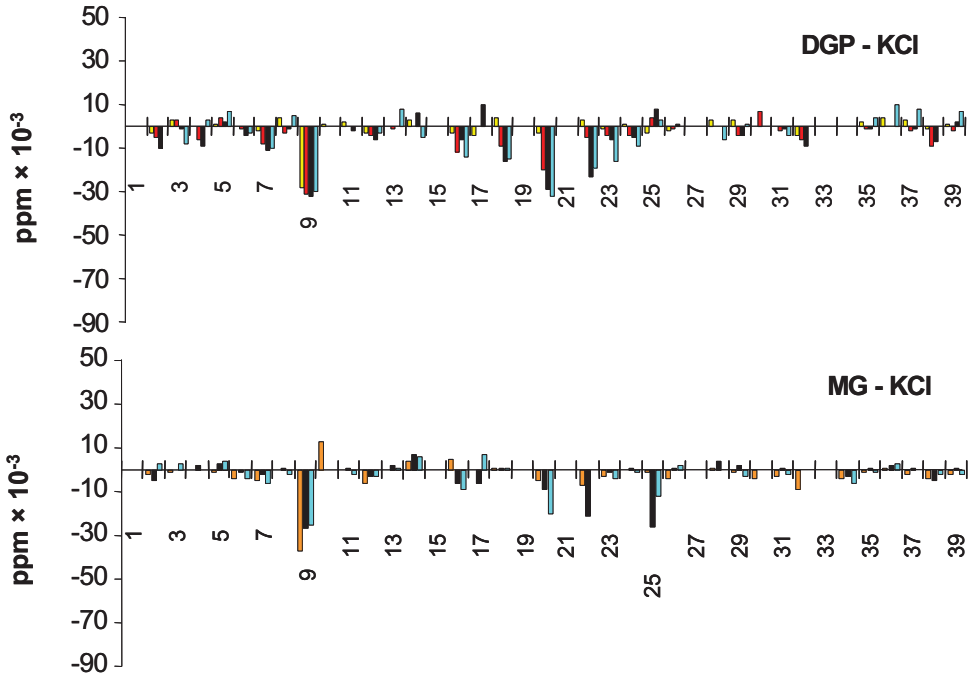


Figure II-5. Variation of the NH chemical shifts for mutant Δ17|29 with the DGP and MG after correction for the ionic strength effect. Chemical shift values obtained with KCl were subtracted from those obtained at the same concentration of DGP and MG. The bars are organised from left to right in increasing solute concentrations with the following colour code: 1 mM (), 2 mM (), 50 mM (), 100 mM () and 200 mM ( ).

Temperature dependence of amide chemical shifts

In general, proton chemical shifts depended linearly with temperature, with the smallest coefficients observed in the metal binding loops (Figure II-6). The metal binding sequences X-Cys-X-X-Cys-Gly-X (X=variable amino acid) are largely conserved among rubredoxins, and comprise residues Val5 to Tyr11, and Ala25 to Ala31 in the Δ17|29 mutant.

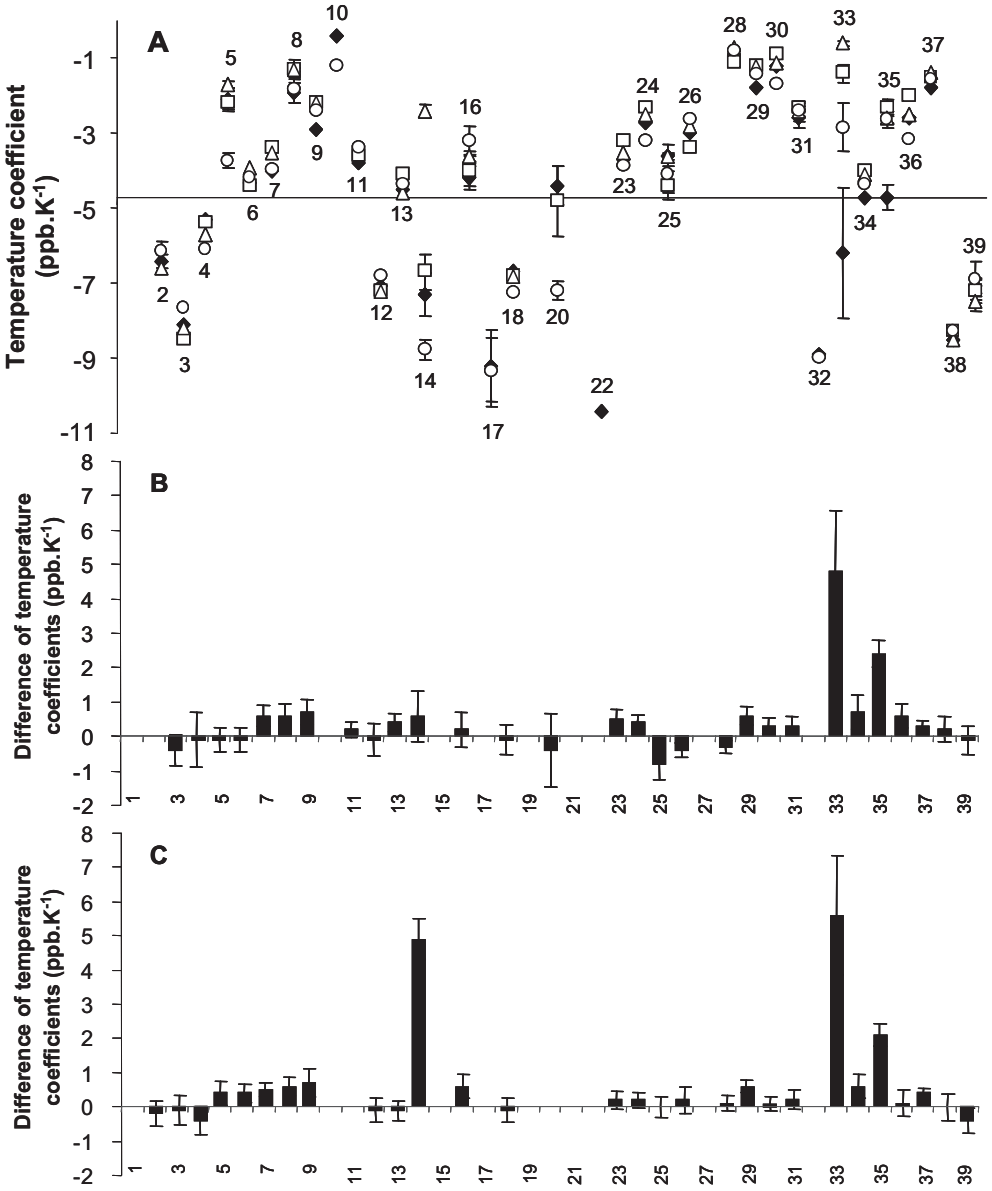


Figure II-6. Temperature coefficients of amide protons for the mutant Δ17|29. (A) Represents the coefficient values for each amide proton in water (◆), 200 mM DGP (□), 200 mM KCl (○) and 400 mM MG (△). The coefficient variation induced by the presence of DGP (B) and MG (C) is also shown. The points on the horizontal axis cover the entire sequence of residues of the protein. The horizontal line at -4.5 ppb·K^{-1} indicates the cut-off value proposed by Baxter and Williamson (1997) for hydrogen-bonded amide protons.

We found a reasonably good correlation between the existence of hydrogen bonds and amide protons with small absolute temperature dependence (values more positive than -4.5 ppb·K^{-1} have been proposed to be a reliable indicator of H-bonding especially if combined with slow exchange rates) (Baxter and Williamson 1997). In the structure of mutant Δ17|29, among the residues with high probability of being involved in H-bonds (determined with the WHATIF software), 79% have temperature coefficients above -4.5 ppb·K^{-1}. The presence of diglycerol phosphate in the sample produced a generally small, but consistent increase in the temperature coefficients of amide protons, with the exception of Lys33 and Ala35, which showed a much larger increase. A similar pattern characterizes the effect of mannosylglycerate except in the case of Asp14 that shows a large increase in the temperature coefficient (threefold) that is not observed in the presence of diglycerol phosphate (Figure II-6).

DISCUSSION

Framework for data interpretation

The aim of this study was to obtain knowledge about the molecular basis of protein stabilization by charged compatible solutes. A series of mutants of RdDg was constructed to investigate the importance of these mutations on the degree of stabilization rendered by solutes. It is well-documented that the thermal denaturation of rubredoxins occurs via a thermodynamically irreversible process (Cavagnero et al. 1998b; Cavagnero et al. 1998a; Eidsness et al. 1997). Therefore, thermodynamic stability parameters are not accessible for these iron-proteins and the stability data reported here refer to kinetic stability, estimated from the half-life for iron release at 90°C.

A link between kinetic stability (half-life) and conformational stability of the native form needs to be established in order to provide a framework for the interpretation of our data. As the precise mechanism of denaturation in these proteins is unknown, that link has to be made on the basis of reasonable assumptions within a model for irreversible denaturation, and the simplest form of the Lumry–Eyring model (Lumry and Eyring 1954) seems appropriate for this purpose. It is reasonable to suppose that the state recently identified in rubredoxins by LeMaster et al. (LeMaster et al. 2005), in which the hydrophobic core is clearly disrupted, represents the unfolded state that denatures irreversibly by loss of the metal centre. Exchange between this state and the native conformation was shown to be fast (LeMaster et al. 2005), in which case a shift in the equilibrium will change the half-life. Because the structure of the metal centre is unchanged in the mutants examined, we may assume that the rate constant for the irreversible step is similar in each case and, hence, the half-life values should correlate with the relative stability of the native and 'unfolded' states. This is far from providing a quantitative relationship between half-life and the intrinsic stability of the native conformations. Nevertheless, it provides the justification for seeking a link between changes in half-life in the presence of solutes and structural features of the native forms

Intrinsic kinetic stability of native rubredoxins and mutants

The deletion of the hairpin loop in RdDg induced a strong decrease in thermal stability. Moreover, the progressive increase in the number of deleted residues (mutants Δ23|29, Δ17|26 and Δ17|29) was accompanied by a progressive decrease of the intrinsic stability, showing that this structural motif is particularly important for the stability of the tertiary structure of rubredoxins, probably by protecting the protein hydrophobic core from solvent access.

Corroborating evidence for this view emerges from the large increase (2.3-fold) in the exposure of the hydrophobic core of mutant Δ17|29 compared with the native structure. It is worth pointing out here that the opening of the middle loop with the concomitant increase in the solvation of the hydrophobic core has been proposed to trigger the loss of the metal ion and subsequent unfolding of *Clostridium pasteurianum* rubredoxin (Bonomi et al. 2000).

The similarity of stability shown by mutant Δ23|29 and the native RdDd, which naturally lacks part of the loop (Table II-1), seems more than a coincidence, and reinforces the positive contribution of this hairpin structure to the stability of this family of proteins. In apparent contrast, the shortening of loops observed in thermophilic proteins compared with mesophilic counterparts has been often proposed as a general strategy for thermostabilization (Auerbach et al. 1998; Bell et al. 2002a). Most likely, this contradiction arises from the fact that the favourable effect of a shorter, rigid hairpin is outweighed, in rubredoxins, by an increase in solvent exposure of the hydrophobic core, with the overall system becoming less stable.

The calculated structure of mutant Δ17|29 revealed notable features, such as a deep cavity in the molecule, and extensive exposure of the aromatic side-chains. The minimal hairpin region of this mutant is responsible for the cavity formation. The structure of the *D. desulfuricans* rubredoxin has also a very short loop but does not show this feature probably because a histidine ring partially fulfils the structural role of the loop (Sieker et al. 1986). It is remarkable that mutant Δ17|29 is able to fold despite the drastic deletion; this reveals the structural importance of the other unaffected motifs in directing folding. Even with a severe disruption of the middle loop, the rest of the characteristic features of the protein structure remain virtually unchanged as

shown by the small RMSD value of 0.82 Å obtained for the superimposition of the mutant and native mean structures.

The progressive decrease in protein stability connected with the shortening of the loop region cannot be ascribed to the size of the loop alone because all the mutants examined, including those with point mutations, showed a clear decrease in their intrinsic stability. In the case of mutant V8N, the exchange of a highly conserved aliphatic side-chain for an uncharged polar group had a striking effect, reducing the half-life for iron release by 65%. Another study reporting mutations on Val8 for more polar residues (V8A and V8D) (Bonomi et al. 2002) demonstrated that the absence of a nonpolar residue at this position dramatically decreased the stability if this rubredoxin. The aliphatic nature of this residue along with three others (5, 38 and 41; sequence numbering of RdDg), as well as their spatial positioning, enables them to pack together, thereby preventing exposure of the metal centre to the solvent. The hydrophobic cluster created by these residues is largely conserved among rubredoxins (Sieker et al. 1994; Blake et al. 1991) and acts like a cap on the tetrahedrally coordinated iron centre, which probably increases its rigidity and compactness, properties generally associated with highly thermostable proteins (D'Amico et al. 2003; Jaenicke 1991; Knapp et al. 1997; Zartler et al. 2001).

The considerable decrease of stability (42%) caused by mutation K17E is probably connected with the addition of an extra negative charge in an already negatively charged patch. The mutation D2K, by contrast, caused only a small decrease of stability (20%). Although this residue does not appear to interact specifically with any other region of the protein (Lamosa et al. 2001), it has been hypothesized that the termini could play a decisive role in the unzipping of the β-sheet (Bougault et al. 2003) and this may explain the observed decrease in stability.

Altogether, the decrease in the intrinsic stability of all tested mutants, even in the case of single-residue mutations, shows that the native conformation of RdDg is remarkably well designed for thermal stability. Moreover, we showed that rubredoxin stability is clearly dependent on the size of the loop region, but it also depends, to a lesser extent, on subtle individual contributions dispersed throughout the protein structure.

Stabilization by compatible solutes

To obtain insight into the mode of action of compatible solutes, we examined the impact of several mutations of RdDg on the degree of stabilization rendered by diglycerol phosphate and mannosylglycerate. In addition, NMR was used to characterize possible interactions of these solutes with the native form of the most perturbed mutant, Δ17|29.

The NMR structure calculation of mutant Δ17|29 has shown that, except for the original loop region, the rest of the protein backbone was virtually unchanged, making it reasonable to assume that the same applies to the other mutants where the deletion was less severe. This assumption is supported by the observation that all mutants retained the UV-visible spectrum displayed by the wild-type Rd, indicating that the metal centre geometry and basic structure was preserved in all the engineered proteins.

Diglycerol phosphate exerted a remarkable stabilization on the wild-type Rd but, surprisingly, was extremely inefficient in the stabilization of the mutants with different size-deletions in the loop region. Given the fact that diglycerol phosphate is a charged solute it is pertinent to analyse the alterations in the electric charge distribution of the loop region associated with the engineering of the loop size. Mutant Δ17|26 has a net charge identical to that of the parent Rd, whereas mutant Δ23|29 shows a decrease of two positive charges and mutant Δ17|29 has a net loss of one positive charge. In the case of

mutant Δ23|29, which is destabilized by the solute, the deletion of seven residues led to the formation of a cluster of four negatively charged residues (DPDSFED), not present in the other mutants that are stabilized by diglycerol phosphate. We hypothesize that the repulsive forces originated from this sequence could contribute to the negative effect exerted by diglycerol phosphate on the stability of this mutant. In agreement with this view, the RdDd, which naturally has a deletion of seven residues in the loop region but lacks this cluster of charged residues, is stabilized by diglycerol phosphate. However, the explanation is surely more complex because mannosylglycerate, which is also negatively charged, does not destabilize mutant Δ23|29, and actually increased its half-life for iron release by 50%. The contrasting behaviour of these equally charged solutes is clear evidence for the distinct nature of the mechanisms underlying protein stabilization by mannosylglycerate and diglycerol phosphate. The differences are not restricted to this mutant. For example, the stability of the wild-type Rd was strongly enhanced by diglycerol phosphate but only modestly improved by mannosylglycerate. In addition, our work demonstrates that minimal alterations in the protein sequence (single mutations) produce considerable differences in the extent of stabilization rendered by a given solute (Figure II-1). Altogether, these results consistently support the view that the effect induced by solutes on protein stability is strongly dependent on the specific protein/compatible solute system examined.

Given the observed specificity of the stabilizing effect, one could hypothesize the existence of specific interactions, or loci for preferential binding on the protein molecule. Proton chemical shifts are very sensitive probes of local fluctuations of the average chemical environment and therefore, were used to look for evidence of preferential interaction sites of the solute with the protein. The pattern of NH shifts induced by the three

charged solutes (DGP, MG and KCl) was broadly similar. However, when the effect of ionic strength was discounted, the differences between diglycerol phosphate and mannosylglycerate became apparent (Figure II-5). The three stabilizing solutes (diglycerol phosphate, mannosylglycerate and trehalose) produce different patterns of chemical shift variation but of similar magnitude, which suggests small, but distinct structural alterations, probably due to specific interactions of the solute with the protein surface. Solutes are generally regarded as causing no major change in protein structure given the low magnitude of chemical shift variations observed in the few studies available (Foord and Leatherbarrow 1998; Lamosa et al. 2003). Although our results corroborate this broad view, we looked for evidence at a much finer level and found some evidence for the presence of small conformational changes. These changes may be large enough to improve the protein stability, and yet, as reflected by the low magnitude of the chemical shifts, too small to affect the overall structure, and probably the physiological function. We should bear in mind, however, that the NMR data was obtained at a temperature lower than the stability data and therefore, the solute/protein interactions could change with temperature.

Preferential sites for solute action are not clearly apparent and probably the interactions are spread throughout the protein surface; however, residues Cys9, Leu20 and Asp22 exhibit shifts that are above the average, these features being common to diglycerol phosphate and mannosylglycerate. Ala25 also experiences a notable shift but only in the presence of mannosylglycerate. Two of these residues (Leu20, Asp22) are located in the poorly structured residual loop and the large effect observed in Cys9 could indicate that this cysteine has the least stable conformation among the iron ligands.

Overall the stabilizing solutes produce mainly negative NH shifts, which is generally associated with stronger hydrogen bonds and therefore a tighter protein structure. The same general effect on chemical shifts (*i.e.*, negative variation) is observed upon lowering the temperature of protein solutions. Further evidence for a more compact structure in the presence of stabilizing solutes is provided by the increase in the temperature coefficients (less negative) of NH groups (Figure II-6). In fact, the signals of amide protons involved in hydrogen bonds generally shift less with temperature (Baxter and Williamson 1997). Therefore, the tendency to increase the coefficients in the presence of the stabilizing solutes is likely to reflect the strengthening of the hydrogen bond network. These findings are in line with an earlier study about the effect of diglycerol phosphate on the dynamics of wild-type RdDg, which reported a restriction of large-scale motions induced by the adition of a stabilizing solute (Lamosa et al. 2003).

In summary, this study provides indication for the existence of at least two effects that could play a role in the complex strategy by which solutes confer higher stability to proteins: an induced overall compaction of the native protein, and specific, weak interactions of the solutes with the protein surface.

ACKNOWLEDGEMENTS AND WORK CONTRIBUTIONS

NMR spectra involved in the structure determination were acquired with the help of Dr. Pedro Lamosa. This work was supported by the European Commission (Contracts QLK3-CT-2000-00640 and COOP-CT-2003-508644 and by Fundação para a Ciência e a Tecnologia (FCT), Portugal, and FEDER (POCTI/BME/35131/2000).

Chapter III

Relationship between protein stabilization and protein rigidification induced by mannosylglycerate

The results of this chapter were published in:
Pais T.M., Lamosa P.L., Garcia-Moreno B., Turner D.L., Santos H. (2009) Relationship between protein stabilization and protein rigidification induced by mannosylglycerate. *J. Mol. Biol.* 394:237-250.

CONTENTS

SUMMARY

Understanding protein stabilization by small organic compounds is a topic of great practical importance. The effect of mannosylglycerate, a charged compatible solute typical of thermophilic microorganisms, on a variant of staphylococcal nuclease (SNase) was investigated using several NMR spectroscopy methods. No structural changes were apparent from the chemical shifts of amide protons. Measurements of ^{15}N relaxation and model-

free analysis, water-amide saturation transfer (CLEANEX-PM), and H/D exchange rates provided a detailed picture of the effects of mannosylglycerate on the backbone dynamics and time-averaged structure of this protein. The widest movements of the protein backbone were significantly constrained in the presence of mannosylglycerate, as indicated by the average five fold decrease of the H/D exchange rates, but the effect on the ms time scale was small. At high frequencies, internal motions of SNase were progressively restricted with increasing concentrations of mannosylglycerate or reduced temperature while the opposite effect was observed with urea (a destabilizing solute). The order parameters showed a strong correlation with the changes in the T_m values induced by different solutes, determined by differential scanning calorimetry. These data show that mannosylglycerate caused a generalised reduction of backbone motions and demonstrate a correlation between protein stabilization and protein rigidification.

INTRODUCTION

The mechanism underlying protein stabilization by osmolytes is a challenging research topic. In an early attempt to explain the effects of stabilizers or denaturants on proteins Tanford and Nozaki proposed that the free energy of transfer of a protein molecule (either in the native or denatured state) from water to an osmolyte solution was given by the sum of the transfer free energies of the solvent-exposed parts. The individual transfer free energies were estimated from the solubility of different peptide model compounds in water and in osmolyte solutions (Nozaki and Tanford 1963; Nozaki and Tanford 1965; Nozaki and Tanford 1970). Later, Timasheff and co-workers showed that, upon unfolding, stabilizing compounds were preferentially excluded from the vicinity of the protein while denaturing agents were

preferentially bound (Arakawa and Timasheff 1985; Timasheff 1992; Timasheff 1993). The protecting effect would then arise from the larger destabilization of the denatured state over the native state. More recently, Bolen and co-workers have established the validity of the transfer free energy model, and concluded that the unfavourable interaction of osmolytes with the protein backbone is the major driving force in protein stabilization (Liu and Bolen 1995; Auton and Bolen 2004; Auton et al. 2007). Yet, despite the significant progress in this area, the molecular mechanism underlying the osmolyte-protein stabilizing interactions remains elusive. Moreover, the relationship between protein stability and dynamics remains obscure.

The terms "osmolytes" and "compatible solutes" have been used indiscriminately in the literature to designate low-molecular mass compounds, either organic or inorganic, that accumulate inside the cell to counterbalance the osmotic pressure of the external medium. However, it has been shown in recent years that the role of these compounds goes beyond osmoprotection. They are involved in the cell response to other types of stress, such as heat or free radicals (Santos et al. 2007). Therefore, throughout this work the terms "compatible solute" or the short form "solute" will be used to discourage the exclusive association of these protecting compounds with osmotic stress.

With the discovery of hyperthermophiles in the early 1980's it became apparent that organisms adapted to hot environments accumulate compatible solutes that are rarely or never found in mesophiles: these organic molecules typically comprise a free carboxyl group or are phosphodiester compounds, hence they are negatively charged, contrasting with the zero net-charge of solutes typically found in mesophiles. Comparative studies on the performance of different solutes have emphasised the superior ability of negatively charged solutes to increase the protein melting temperature (T_m). For example, glycerol and trehalose increased the T_m of staphylococcal nuclease (SNase) by

0.8°C and 12°C per molar, respectively, while mannosylglycerate and mannosyl-lactate induced increments of 17°C and 22°C per molar on the same protein (Faria et al. 2008). However, our attempts to rationalise the relative magnitudes of the stabilization induced by structurally related solutes on different enzymes were largely frustrated. Even more surprisingly, the degree of stabilization rendered by a particular solute on a series of single mutant proteins varied significantly (Pais et al. 2005). These results are difficult to explain with any of the models proposed for the stabilization of proteins by solutes. Putative specific interactions with the surface of the native protein might account for the observed effects, but their existence lacks clear evidence.

Numerous factors that can contribute to the intrinsic stability of globular proteins, especially those from hyper/thermophilic organisms have been put forward (Razvi and Scholtz 2006). Hydrophobic packing, hydrogen bond networks, salt bridges, and Coulomb interactions can increase the stability of a protein if properly optimized (Petsko 2001). Thus, it is possible that the mode of action of stabilizing solutes involves the optimization of at least some of these features. This line of thought leads us to the central question of how can solutes alter such properties.

It is generally accepted that the structures of proteins are not affected significantly by the presence of solutes, and that structural changes probably do not play a major role in stabilization. On the other hand, clear alterations in protein dynamics have been observed by us and by others (Doan-Nguyen and Loria 2007; Lamosa et al. 2003; Wang et al. 1995). Therefore, a comprehensive NMR study of the changes induced by solutes was performed to address the following questions. Is there a correlation between stabilization by solutes and rigidification of the protein? Is this rigidification a global event

or is it selective for specific regions of the protein? Is stabilization associated with motional restrictions at specific dynamic regimes?

Protein dynamics comprises a wide range of motions, from small vibrational fluctuations of bond lengths (ps-ns time scale) to translations of atom groups (micro-millisecond time frame) to concerted motions of whole structural elements (second to minute time scale). For this reason, it is important to access a wide range of motions and time scales to characterise protein dynamics. We used NMR spectroscopy and performed spin-relaxation measurements to study fast protein motions at the ps-ns time scale, magnetization transfer experiments (with the CLEANEX-PM sequence) to assess exchange on the ms time scale, and conventional proton/deuterium amide exchange measurements to probe events in the minute time frame. The stabilizing solute mannosylglycerate (potassium form) typical of hyperthermophilic organisms and a hyperstable variant of staphylococcal nuclease with three amino acid substitutions (P117G,H124L,S128A) were selected as the model system for this work. KCl and glycerol were used as controls for ionic strength and viscosity, respectively. Changes in protein dynamics were correlated with changes in the melting temperature of the model protein as determined by calorimetric measurements.

MATERIALS AND METHODS

Protein production

The PHS variant of SNase protein includes three substitutions: P117G, H124L and S128A. This protein is considerably more stable than the wild type (Chen et al. 2000). The plasmid containing the gene for PHS SNase was transformed in *Escherichia coli* BL21(DE) for over-expression. Cells were grown at 37°C on ^{15}N-labelled defined medium (Lamosa et al. 2003) (or LB medium)

supplemented with kanamycin. Protein over-expression was started by adding β-D-thiogalactopyranoside (1 mM final concentration) to the cell culture at an OD_{600nm} of 0.6 units. The cells were harvested five hours after the induction by low speed centrifugation (2000 × *g*, 10 min). Protein purification proceeded as described elsewhere (Shortle and Meeker 1989) with some modifications: the urea buffer was adjusted to pH 8, the protein solution was loaded onto a single column (Sepharose Fast Flow, Pharmacia Uppsala Sweden), acetone was not used for the precipitation steps, and the final dialysis was carried out first in 1 M KCl and then in distilled H_2O.

The final protein concentration was calculated by measuring the absorbance at 280 nm and using an extinction coefficient of 0.93 ($cm \cdot mg \cdot ml^{-1}$). Purity was above 95% as assessed by SDS-PAGE with coomassie staining. The purified protein was flash-frozen with liquid nitrogen in small drops and stored at -80°C.

Differential scanning calorimetry

Differential scanning calorimetry (DSC) was performed on a MicroCal VP-DSC MicroCalorimeter equipped with 0.51 ml cells and controlled by the VP-viewer program (Microcal, Century City, CA, USA). Temperature and heat flow were calibrated according to MicroCal instructions. Stock solutions were prepared by extensive dialysis of concentrated SNase against phosphate buffer (10 mM, pH 7.5). A volume of 2 ml of phosphate buffer with or without solute was prepared and divided equally into two Eppendorf tubes. An aliquot of the concentrated protein solution or an equal volume of phosphate buffer was added to each of the tubes. These solutions were then used to fill the sample and reference cells, respectively. DSC scans were run from 20 to 90°C at a constant heating rate of 1°C/min. Reversibility was assessed by performing two sequential DSC scans with the same protein solution. The transition

temperature was determined, in each case, on the basis of three independent runs. The tested solutes were mannosylglycerate (0.15, 0.25, 0.35 and 0.50 M), glycerol (0.60 M), urea (0.25 M) and KCl (0.25 M).

NMR spectroscopy

Water-amide proton saturation transfer experiments (CLEANEX-PM) were acquired on a Bruker AVANCE^{III} 800 spectrometer (Bruker, Rheinstetten, Germany) operating at 800.33 MHz, using a 5 mm inverse detection probe with field gradients on the z-axis. All other NMR experiments were performed on a Bruker DRX 500 spectrometer (Bruker, Rheinstetten, Germany) at 500.13 MHz, with a broadband inverse detection 5 mm probe head with triple axis pulsed field gradients. All NMR data were collected using ^{1}H-^{15}N HSQC detection schemes. The control sample used for the NMR experiments contained uniformly ^{15}N-labeled SNase (2.6 mM), and acetate-d_4 buffer (60 mM, pH 5.1), except for the CLEANEX-PM experiments where phosphate buffer (10 mM, pH 7.3) was used instead. Other samples differ from the control by the presence of one of these solutes: mannosylglycerate (0.15 M, 0.25 M and 0.35 M), glycerol (0.60 M), urea (0.25 M) or KCl (0.25 M). Glycerol and KCl were used to assess, respectively, the contributions of viscosity and ionic strength to the results observed with 0.25 M mannosylglycerate. The viscosity of 0.6 M glycerol is identical to that of 0.25 M mannosylglycerate. The viscosity of mannosylglycerate solutions was determined using an Ubbelohde type viscometer.

The assignment of the resonances present in the HSQC spectrum was performed by comparison with previously published data on the SNase structure (Wang et al. 1992a; Wang et al. 1992b). The resonances corresponding to the mutated residues were not assigned, since NMR information was not available at the time of this study.

^{15}N relaxation measurements and analysis

^{15}N relaxation rates were measured for the amide groups of SNase at 37°C in the absence (control) and in the presence of four solutes: mannosylglycerate (0.15, 0.25 and 0.35 M), urea (0.25 M), glycerol (0.60 M), and KCl (0.25 M). Studies with higher concentrations of mannosylglycerate were hampered by severe line broadening. Two additional sets of relaxation measurements were carried out for the control sample (no solutes) at 32°C and 42°C. Longitudinal (T_1) and transverse (T_2) relaxation times, and ^{1}H-^{15}N steady state heteronuclear NOE (h-NOE) values were measured using conventional sequences (Kay et al. 1989) with some modifications regarding water suppression. For T_1 measurements, a 3-9-19 WATERGATE sequence was used (Piotto et al. 1992), while in the case of T_2, water suppression was achieved by a WATERGATE sequence with selective inversion of the water signal. In the h-NOE experiment without proton saturation, water suppression was achieved by selective inversion of the water resonance and by the use of gradient pulses to select for the ^{15}N → ^{1}H coherence transfer pathway.

T_1 relaxation rates (R_1) were obtained with relaxation periods of 0, 20, 50, 100, 250, 500, 1000, 1250, and 1500 ms. Values for the T_2 relaxation rates (R_2) were obtained with a Carr-Purcell-Meiboom-Gill (CPMG) spin-echo period of 500 μs, and relaxation periods of 0, 16, 32, 48, 56, 64, 96, and 144 ms. In order to measure h-NOEs, spectra acquired with and without saturation of protons were recorded in an interleaved manner to minimize systematic differences. ^{1}H saturation was achieved by the application of 120º pulses spaced at 5 ms intervals for 3 seconds (Markley et al. 1971). The total recycling delay for both experiments was 4.2 seconds.

The number of scans accumulated per t_1 increment for the T_1, T_2, and h-NOE experiments was 24, 32, and 36 scans, respectively. The peak intensities were measured using XWinNMR software (Bruker, Rheinstetten,

Germany). The h-NOE values were calculated using the ratio between the volume of each ^{1}H-^{15}N correlation peak in the presence or absence of proton saturation. Relaxation rates for T_1 and T_2 were determined by non-linear least-squares fitting of the peak volumes to a mono-exponential decay function. Error estimates were obtained from the standard deviation of the fitting coefficients.

The Tensor2 (Dosset et al. 2000; Cordier et al. 1998) program was used to characterise the diffusion tensor of the SNase by analysing the R_2/R_1 ratios obtained for each of the examined conditions. Residues with significant internal motion or in fast chemical exchange were excluded from the characterisation of the diffusion properties of the protein (Tjandra et al. 1995). Error estimates were obtained from 600 Monte Carlo simulations while chi-square and *F*-test analysis were used to evaluate the quality of the fit and the statistical significance of introducing an extra parameter to the diffusion model (isotropic, axial, or fully anisotropic diffusion). The same program was used to select the dynamic model that best described the internal mobility of the protein backbone based on the relaxation data collected. This software uses the Lipari-Szabo type analysis (Lipari and Szabo 1982a; Lipari and Szabo 1982b). The analysis was performed using an average amide bond length of 1.02 Å and a chemical shift anisotropy value approximated to -170 ppm for ^{15}N nuclei (Tjandra et al. 1996).

Five increasingly complex models were tested iteratively as described by Mandel et al. (1995) (Mandel et al. 1995) optimizing the parameters S^2, τ_e, R_{ex}, S^2_f, and S^2_s, where S^2 is the generalised order parameter characterising the amplitude of the internal motions, τ_e is the effective correlation time for the internal motions, R_{ex} is the exchange contribution to T_2, and subscripts *f* and *s* indicate fast and slow time scale motions, respectively.

Model selection was done automatically by the program Tensor2 and error estimates obtained by 600 Monte Carlo simulations. This procedure was repeated with the data collected from all the other conditions at which ^{15}N relaxation experiments were performed.

Water-amide proton saturation transfer experiments

Water-amide proton exchange rates in the millisecond range were measured by saturation transfer spectroscopy using the phase-modulated CLEAN chemical exchange technique (CLEANEX-PM) with a fast-HSQC detection scheme (Hwang et al. 1997; Hwang et al. 1998; Mori et al. 1995). This technique has been shown to suppress NOE contributions from other protons with chemical shifts coincident with water (*e.g.,* $C^{\alpha}H$) as well as contributions from exchange-relayed NOEs of rapidly exchanging protons. Thus, it was possible to measure pure exchange rates for the observed amide groups. The experiments were performed at 37°C on a Bruker AVANCE^{III} 800 spectrometer. Four samples of SNase were prepared in 10 mM phosphate buffer, pH 7.3, without solutes or with one of the following additives: mannosylglycerate (0.25 M), KCl (0.25 M), and glycerol (0.60 M). As stated already, glycerol and KCl were used, respectively, as controls for the viscosity and ionic strength of mannosylglycerate (0.25 M).

The CLEANEX-PM experiments were collected with mixing times of 5, 8, 10, 15, 20 and 35 ms with repetition of the 20 ms mixing time experiment to estimate the standard deviation of peak volumes. These spectra were acquired with 2048 complex data points, 256 time increments, and 8 scans per increment. A delay of 3 seconds was used between scans to minimize the level of water saturation.

Additional NMR experiments were necessary to quantify water-amide exchange rates accurately (Hwang et al. 1998). These included (i) fast-HSQC

spectra for determination of reference peak volumes, (ii) measurements of water saturation level in the CLEANEX-PM sequence, and (iii) inversion recovery spectra for water T_1 determination, according to the method described by Hwang et al. (1997) (Hwang et al. 1997).

Proton/Deuterium exchange rates

Proton/Deuterium exchange experiments were performed to monitor exchange rates of slowly exchanging amide protons (min to hour time scale), using protein samples with KCl (0.25 M), mannosylglycerate (0.25 M) or glycerol (0.60 M). Exchange rates were measured for the protein samples at six temperatures: 30, 33, 37, 42, 44 and 47°C. Glycerol containing samples were only examined at 37°C. The proton/deuterium exchange reaction was initiated by dissolving the lyophilised protein samples in 99.9% ^2H_2O. The resulting pH of 5.7 was measured with a glass electrode without further corrections.

The exchange rates were determined by measuring the volume of amide cross-peaks in HSQC spectra as a function of time. Spectra were acquired with 2048 complex data points, 256 time increments, and 8 scans per increment. All experiments started less than 25 minutes after the exchange reaction was initiated. Standard Bruker pulse sequences were used to collect HSQC spectra with sensitivity enhancement. The peak volumes were measured with XWinNMR software (Bruker, Rheinstetten, Germany) and fitted to a first order decay curve by nonlinear regression, providing the exchange rate for each amide proton.

Amide-proton chemical shift variation

Chemical shift variation as a function of solute concentration was followed at 37°C by ^1H-^{15}N HSQC spectra with the same acquisition parameters as in the H/D exchange experiments. Increasing amounts of KCl or mannosylglycerate

(up to 0.25 M final concentration) were added to an initial control sample (no solutes) and HSQC spectra were recorded for each condition. Protein samples containing either 0.25 M urea or 0.60 M glycerol were also examined by ^{1}H-^{15}N HSQC spectra.

The temperature dependence of amide proton chemical shifts was also evaluated for SNase samples in the absence of solutes and in the presence of 0.35 M mannosylglycerate. A temperature range of twenty two degrees (27°C to 49°C) was used to determine the amide proton temperature coefficients of all accessible resonances. The chemical shifts were measured using the TOPSPIN 2.0 software (Bruker, Rheinstetten, Germany) and referenced to dioxane in a capillary.

RESULTS

Differential scanning calorimetry

The unfolding transition temperatures (T_m) of SNase determined in the presence of urea, glycerol, KCl and mannosylglycerate are shown in Table III-1. Glycerol and KCl were used to assess, respectively, the contributions of viscosity and ionic strength; the viscosity of 0.6 M glycerol is identical to that of 0.25 M mannosylglycerate. In the absence of solutes, the T_m was 67.4°C. Mannosylglycerate increased the T_m in a concentration dependent manner. On the other hand, 0.60 M glycerol and 0.25 M KCl had little effect, and 0.25 M urea decreased the T_m by 1.6°C. Furthermore, the presence of mannosylglycerate did not affect the extent of reversibility of the protein unfolding reaction.

Table III-1. Transition temperatures for the unfolding of SNase obtained by DSC.

	Conc. (M)	T_m (°C)	ΔT_m[a] (°C)
No solute	0	67.40 ±0.41	0
Urea	0.25	65.80 ±0.23	- 1.60
KCl	0.25	66.64 ±0.10	- 0.76
Glycerol	0.60	67.21 ±0.22	- 0.19
Mannosylglycerate	0.15	69.15 ±0.29	+ 1.75
	0.25	70.58 ±0.11	+ 3.18
	0.35	72.00 ±0.15	+ 4.60
	0.50	74.28 ±0.20	+ 6.88

[a]Difference of SNase T_m value in the presence and in the absence of different solutes.

Relaxation data

Transverse relaxation rates (R_2) were determined as a function of protein concentration using samples with 0.5 mM, 1.0 mM, 1.5 mM, 2.5 mM and 3.5 mM of SNase in the presence of 0.35 M mannosylglycerate to test for possible protein aggregation (not shown). These results show that there is no significant concentration dependence of R_2 below 3.5 mM and a protein concentration of 2.6 mM was used for the remaining experiments.

Longitudinal (R_1) and transverse relaxation rates as well as h-NOE values were obtained at the different experimental conditions for a total of 93 backbone amide groups, accounting for 80% of the backbone NHs previously assigned (Wang et al. 1992b; Wang et al. 1992a). The average relaxation rates and h-NOE values measured in the control conditions (absence of solutes) are in general agreement with those obtained for another SNase variant (H124L) (Alexandrescu et al. 1996).

With few exceptions, the calculated R_2/R_1 ratios were very homogeneous for the protein backbone in each of the conditions tested (Figure III-1).

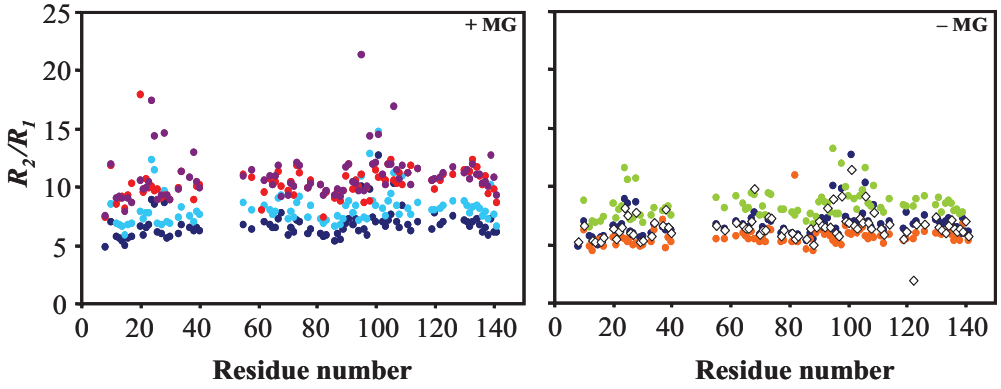


Figure III-1. Effect of mannosylglycerate and other solutes on the R_2/R_1 ratio of each amide group. The left-hand panel illustrates the variation of R_2/R_1 ratios as a function of mannosylglycerate concentration: 0 M (dark blue circles), 0.15 M (light blue circles), 0.25 M (red circles), and 0.35 M (purple circles). The panel on the right shows the R_2/R_1 ratio of amide groups in the absence (dark blue circles) and in the presence of 0.25 M KCl (orange circles), 0.25 M urea (empty diamonds), and 0.60 M glycerol (green circles). MG stands for mannosylglycerate.

The total correlation time (τ_c) of the protein was estimated from the average R_2/R_1 ratio, excluding values that fail the selection criteria described by Tjandra et al. (1995). This correlation time showed a positive dependence on mannosylglycerate concentration, while the opposite effect was observed with KCl and urea, which appears to be in fair agreement with the variation of the viscosity of the solution. However, bulk viscosity was not solely responsible for the estimated total correlation time since glycerol (0.6 M) gave rise to a τ_c of 8.4 ns while mannosylglycerate, at the same bulk viscosity (0.25 M), gave rise to a τ_c of 10.4 ns (Table III-2). In the absence of solutes an increase in the experimental temperature caused a decrease in the total correlation time. The h-NOE values showed no apparent trend as a function of mannosylglycerate concentration.

Table III-2. Characterization of the diffusion tensor obtained for SNase at different experimental conditions, using the Tensor2 program (Dosset et al. 2000).

	32°C	37°C							42°C
	NS	**Urea**	**NS**	**KCl**	**Gly**	**MG 0.15 M**	**MG 0.25 M**	**MG 0.35 M**	**NS**
τ_c **(ns)**[a]	8.6 ±1.1	6.5 ±1.0	6.8 ±1.2	5.6 ±1.0	8.4 ±1.3	8.7 ±1.8	10.4 ±1.8	10.3 ±1.5	6.2 ±1.2
α **(deg)**[b,c]	-84.9	-76.2	-70.9	-78.9	-66.9	-61.5	-72.3	-52.1	-67.0
	±14.1	±8.1	±7.9	±10.2	±6.5	±14.9	±7.9	±8.5	±12.2
β **(deg)**	-64.6	-65.6	-64.7	-64.0	-64.6	-63.8	-65.4	-68.6	-63.5
	±3.3	±1.6	±1.5	±1.8	±1.7	±2.0	±1.5	±2.7	±2.0
γ **(deg)**	-20.7	-18.7	-16.0	-7.2	-19.0	-14.0	-9.3	-21.3	-16.7
	±5.2	±2.7	±2.5	±3.0	±2.6	±3.0	±2.5	±3.6	±3.2
D_x **10^7 s^{-1}**	1.40	1.67	1.65	1.80	1.43	1.46	1.25	1.28	1.75
	±0.03	±0.02	±0.02	±0.02	±0.02	±0.02	±0.01	±0.02	±0.03
D_y **10^7 s^{-1}**	1.58	1.86	1.83	1.98	1.61	1.56	1.39	1.44	1.92
	±0.04	±0.02	±0.02	±0.03	±0.02	±0.02	±0.01	±0.02	±0.03
D_z **10^7 s^{-1}**	1.90	2.34	2.30	2.48	1.95	1.92	1.68	1.70	1.45
	±0.03	±0.02	±0.02	±0.03	±0.02	±0.02	±0.01	±0.02	±0.03

[a] The total correlation time (τ_c) was calculated using the trimmed average of R_2/R_1 ratios, with errors taken from the standard deviation, according to the procedures outlined by Tjandra et al. (1995). NS stands for no solute, Gly for glycerol and MG for mannosylglycerate. [b] α, β, γ, D_x, D_y, D_z describe the orientation and amplitude of the principal components of the diffusion tensor in the frame of the PDB structure 1EY8. [c] The errors were estimated on the basis of 600 Monte Carlo simulations using a routine embedded in the Tensor2 program.

Diffusion tensor

The diffusion tensor and all subsequent parameters of the internal dynamics were obtained from the experimental data using the crystal structure of the PHS variant of SNase (PDB entry code: 1EY8) as a structural model (Figure III-2). This structure is markedly asymmetric, as shown by the principal components of the inertial tensor (I_{xx}:I_{yy}:I_{zz} = 1.00:0.94:0.71). The rotational

diffusion tensor of the SNase was fitted for each examined condition using relaxation data from 38 residues that passed the selection criteria outlined by Tjandra et al. (1995) in all cases. The statistical analysis routine embedded in the program Tensor2 (Dosset et al. 2000; Cordier et al. 1998) was used to test models for the diffusion tensor of the protein; the F-test showed that the fully anisotropic model gave a statistically significant improvement over the isotropic or the axially symmetric models in each case (Fexp > 0.1 Ftest). The data presented in Table III-2 characterises the diffusion tensor of the protein for all tested conditions.

The orientation of the diffusion tensor relative to the inertial tensor did not change significantly in the presence of the four examined solutes or with temperature. The diffusion rates (D_x, D_y, D_z) were slowed down considerably in the presence of increasing concentrations of mannosylglycerate. The addition of glycerol also caused a reduction in the diffusion rates, but to a lesser extent than mannosylglycerate (0.25 M), while KCl and urea had the opposite effect. Not surprisingly, increasing temperatures also gave rise to increasing diffusion rates.

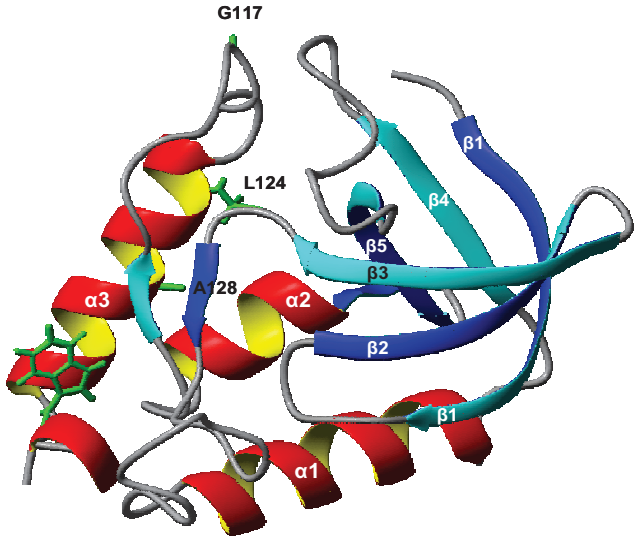


Figure III-2. Ribbon diagram of the PHS variant of staphylococcal nuclease. Elements of secondary structure are labelled. Views were generated with MolMol (Koradi et al. 1996) using PDB entry 1EY8.

Internal mobility

The internal mobility of each NH vector was characterised by one of 5 possible models (see Material and Methods) according to the flow-chart proposed by Mandel et al. 1995. In each case, the majority were fitted by the simplest model (Table III-3). The addition of glycerol and mannosylglycerate did not appear to change significantly the nature of the mobility of the amide groups. However, KCl induced a considerable simplification of the dynamics (22 extra residues were fitted with the simplest model), mostly due to reduced chemical exchange contributions (models 3 and 4).

Table III-3. Number of residues assigned to each dynamic model at different experimental conditions.

	32°C	37°C							42°C
Model [a]	**NS**	**Urea**	**NS**	**KCl**	**Gly**	**MG 0.15 M**	**MG 0.25 M**	**MG 0.35 M**	**NS**
1 (S^2)	72	54	51	73	54	64	58	48	57
2 (S^2, τ_c)	6	4	6	2	2	2	4	5	13
3 (S^2, R_{ex})	9	17	19	10	24	15	10	20	11
4 (S^2, τ_c, R_{ex})	4	3	6	1	1	2	2	3	4
5 (S^2_s, S^2_f, τ_c)	2	13	11	7	9	10	18	16	8
Not fitted	0	2	0	0	1	0	1	1	0

[a] S^2 is the square of the generalized order parameter characterizing the amplitude of the internal motions; τ_c is the effective correlation time for the internal motions; R_{ex}, is the exchange contribution to T_2, and the subscripts *f* and *s* indicate fast and slow time scales, respectively. NS stands for no solute, Gly stands for glycerol and MG stands for mannosylglycerate.

The internal mobility of this protein on the ps-ns time scale is intrinsically small with an average S^2 value of 0.896 in the absence of stabilizing solutes at 37°C; among the residues that were analysed only Gly50 and Ala69 have S^2 values below 0.80. Still, the presence of mannosylglycerate induced a further restriction on the fast-scale movements of the backbone amides (Figure III-3). The magnitude of this restriction, reflected by the order parameter average, displayed an almost linear correlation with the increase of the melting temperature caused by the addition of mannosylglycerate. This correlation can be generalized by plotting all of the data as a function of (T_m-T), in which T_m is the melting temperature of unfolding as measured by DSC and T is the experimental temperature (Figure III-4). Solutions containing glycerol and KCl, as controls for the effects of viscosity and ionic strength comparable with 0.25 M mannosylglycerate, had a much smaller influence on S^2 values than

mannosylglycerate. Urea, a destabilizing solute, brought about a decrease in the average S^2 values.

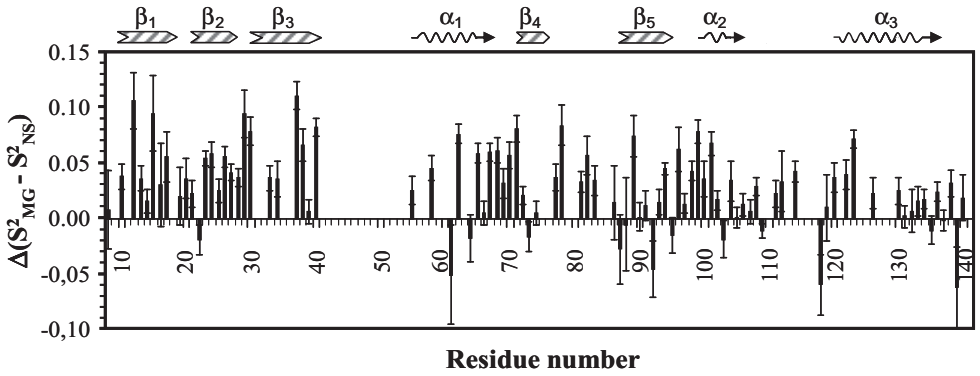


Figure III-3. Changes in the generalised order parameters (S^2) induced by the presence of 0.35 M mannosylglycerate. Subscripts MG and NS refer to 0.35 M mannosylglycerate and no solute addition, respectively. Elements of secondary structure are also represented.

There was no obvious correlation between S^2 values of individual residues and residue solvent exposure, side-chain or surface charge. However, secondary structural elements appeared to slightly influence the effect of mannosylglycerate on the mobility of amide bonds (the S^2 values averaged over each secondary structural element of the protein are shown in Figure III-5). Further support for this observation is obtained by analyzing the average slope for the S^2 variation as a function of mannosylglycerate concentration: this is 30% larger for the regions of defined secondary structure (0.10 M^{-1}) than for the regions of ill defined structure (0.07 M^{-1}). The motions of the β-strands appear to be more restricted by the presence of mannosylglycerate than those of α-helices.

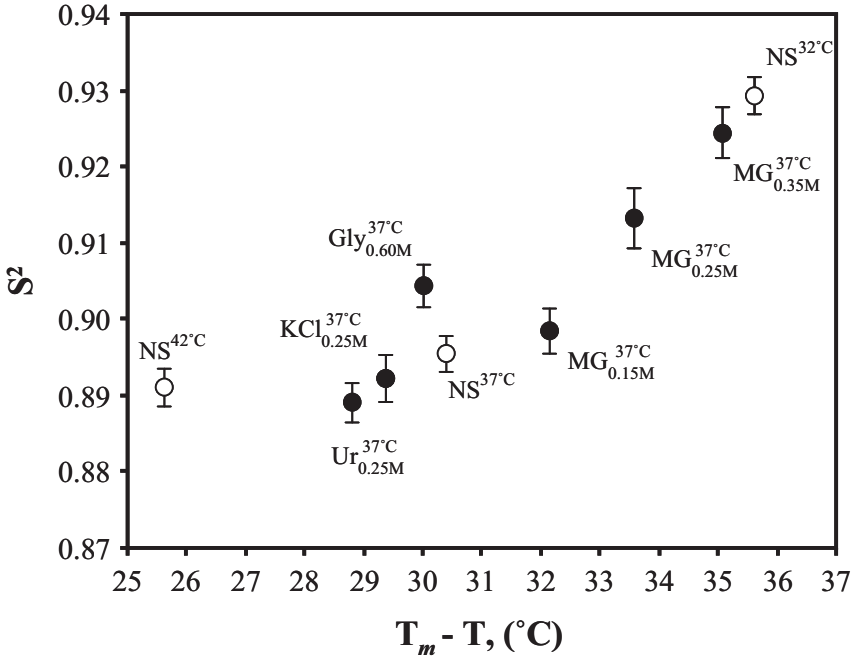


Figure III-4. Plot of the average S^2 values as a function of the shift away from the melting temperature. The (T_m-T) parameter is the difference between the temperature at which relaxation values were measured and the T_m of the protein in the presence of a given solute; NS (no solute), Gly (glycerol), KCl, Ur (urea), MG (mannosylglycerate). The subscripts give the concentration of the solute (M) and the superscripts are the experimental temperatures (°C). The errors bars represent the error of the average calculated from the individual S^2 uncertainties.

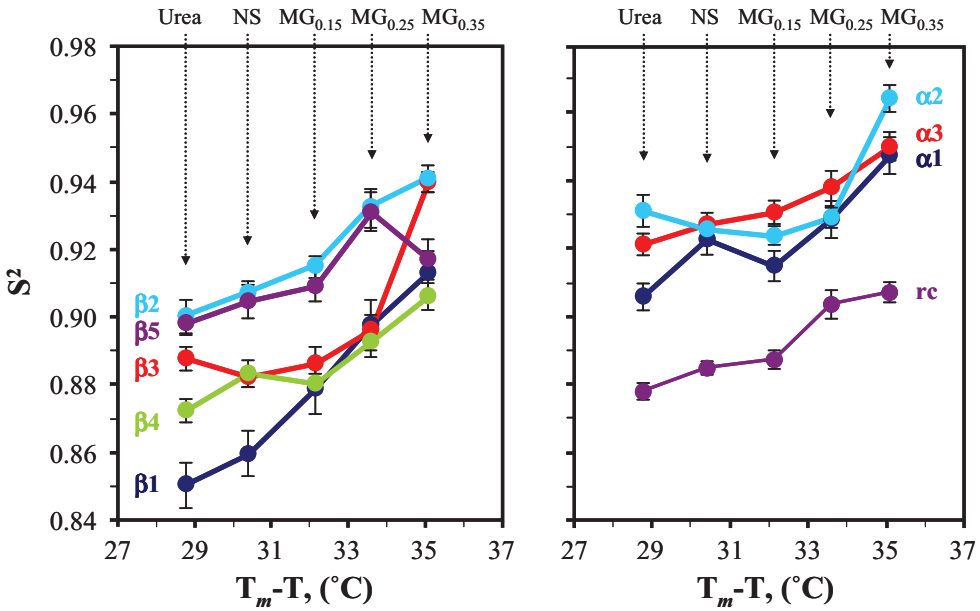


Figure III-5. Average S^2 values of the secondary structural elements of SNase as a function of the shift away from the melting temperature. The amide protons were grouped as a function of secondary structural elements and the respective S^2 values were averaged. The different experimental conditions are specified on top of the plots with NS standing for no solute, and MG for mannosylglycerate (subscripts indicate solute concentration in molar units)

Water-amide proton saturation transfer experiments

In the CLEANEX-PM experiment (Hwang et al. 1997), the dependence of peak volumes on mixing time (τ_m) is given by the following equation (Jeener et al. 1979; Schwartz and Cutnell 1983; Dobson et al. 1986; Mori et al. 1996):

$$\frac{V}{V_0} = [k\,/(R_1 a + k - R_1 b)] \times \{\exp(-R_1 b \times \tau_m) - \exp[-(R_1 a + k)\tau_m]\}$$

[III.1]

where k is the normalized rate constant related to the pseudo-first-order forward rate constant $k_{AB}(NH \rightarrow H_2O) = X_B \cdot k$, with X_B being the molar fraction of

water (≈ 1), V the peak volume obtained from the (CLEANEX-PM)-FHSQC spectra at a given mixing time, τ_m, and V_0 the reference peak volume obtained with the FHSQC spectrum. The R_1a and R_1b parameters refer to the longitudinal relaxation rates of the amide and of the water molecule, respectively. At very short mixing times, equation III.1 can be approximated by:

$$\frac{V}{V_0} = k \times \tau_m \qquad \textbf{[III.2]}$$

and the water-amide exchange rates can be calculated from the initial slope of the plot V/V_0 vs τ_m. The degree of water saturation in each experiment was determined and used to correct the values obtained from the initial slope analysis (Hwang et al. 1998) simply by dividing *k* by the fraction of unsaturated water.

In the absence of solutes and at 37°C, a 35 ms CLEANEX-PM mixing time allowed the observation of 18 assigned resonances: H8, A17, D19, G20, K28, G29, T33, N68, K70, T82, G86, G96, K97, M98, A102, N119, T120, and Q123. Some of these resonances (D19, G20 N68, K70 T82, G86, A102, and T120) were not detected in the CLEANEX-PM spectra with shorter mixing times, so their exchange rates were not calculated. Resonances of L14, G55, and K70 were additionally detected in the presence of KCl, but only with the 35 ms mixing time.

The water-amide proton exchange rates of all accessible resonances except K97 were slower in the presence of glycerol and mannosylglycerate (Figure III-6). Mannosylglycerate induced a marginally larger reduction of the exchange rates than glycerol. KCl induced only a slight decrease in the magnitude of the exchange rates. On average, mannosylglycerate induced a 1.8 fold decrease in the exchange rates while glycerol and KCl caused,

respectively, an average reduction of 1.5 and 1.2 fold. The ratio between exchange rates in the presence and in the absence of solutes was similar for all accessible residues, with a standard deviation of 17% in the case of mannosylglycerate, and 12% in the case of glycerol and KCl. There is no clear correlation between these effects and the stability of SNase.

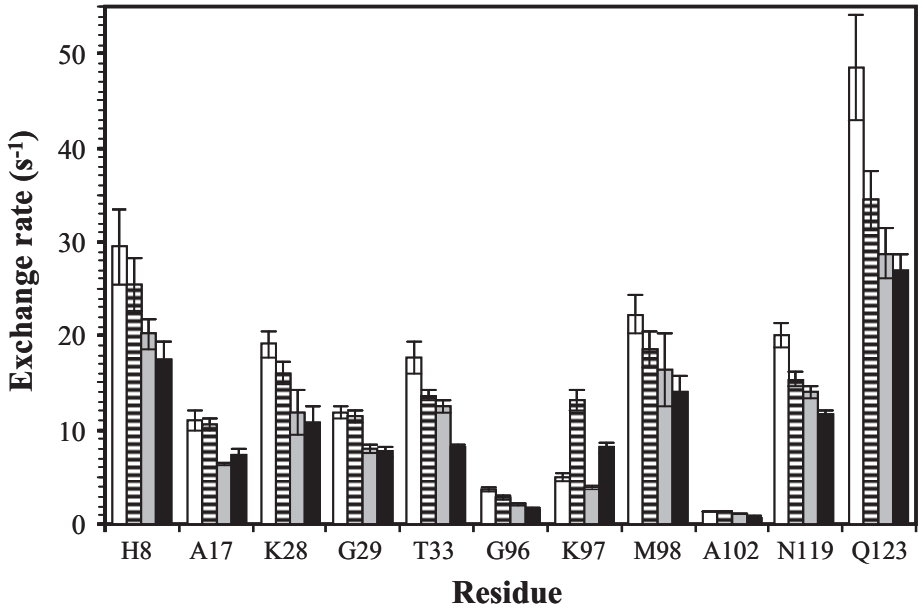


Figure III-6. Fast water-amide proton exchange rates (in s^{-1}) obtained for SNase using the CLEANEX-PM technique. The plot depicts amide protons that exchange with the solvent in the millisecond range at 37°C and pH 7.3 (10 mM phosphate buffer). Rates measured in the absence of solutes (empty bars), in the presence of 0.25 M KCl (striped bars), 0.60 M glycerol (grey bars) and 0.25 M mannosylglycerate (solid bars).

Hydrogen/deuterium exchange rates

The amide protons of the PHS variant of SNase exchange with deuterium from the solvent at widely different rates, from milliseconds to several hours. Amide protons with life-times shorter than the experimental dead time of 25 minutes could not be observed by this method. In the presence of KCl (0.25 M) at 37°C the exchange rates of 66 amide protons could be determined; 22 other

amides were detected only in the first spectrum, having life-times close to the dead time, and one other, V99, exchanged extremely slowly. With mannosylglycerate, five extra resonances were detected in the first spectrum but the rates of exchange were too fast to be measured. The H/D exchange rates displayed a general decrease in samples containing mannosylglycerate, with a maximum 38 fold for V66 at 37°C (Figure III-7). In the presence of glycerol (viscosity control) a decrease of the exchange rates (as compared to KCl) was also observed but it was much smaller than with mannosylglycerate.
Solute induced changes showed no obvious correlation with properties such as accessible molecular surface, residue charge or hydrophobicity. Nevertheless, a slight trend was observed with the effect of mannosylglycerate being more pronounced for the most protected amide protons (Figure III.S-1).

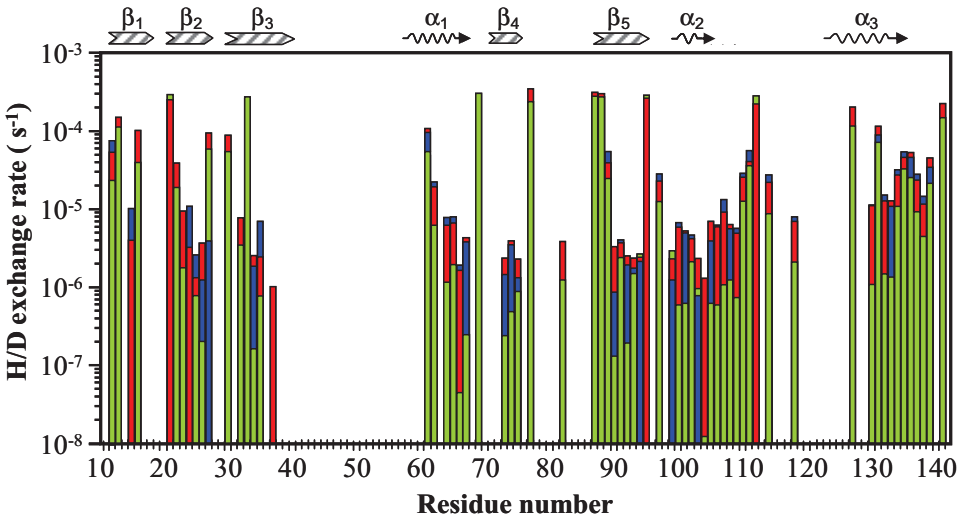


Figure III-7. Effect of mannosylglycerate on the amide H/D exchange rates. The plot illustrates the exchange rates of amide protons of SNase at 37°C in the presence of 0.25 M of mannosylglycerate (green bars), 0.25 M of KCl (red bars) and 0.60 M glycerol (blue bars). The rates are shown on a logarithmic scale and the horizontal axis represents the amino acid sequence. For each residue, the bars are ordered according to the magnitudes of the exchange rates, with the lowest exchange rate at the front, and overlaid. Thus, the dominant green colour shows that the lowest rates generally occur in the presence of mannosylglycerate.

According to the scheme shown in equation III.3, the hydrogen exchange rates can provide information on the thermodynamics of the structural opening reaction that allows the H/D exchange (Englander and Mayne 1992). In stable folded proteins, the so called *EX2* regime holds ($k_{cl} >> k_{ch}$) (Hvidt and Nielsen 1966; Englander et al. 1996) and equation III.4 can be used to estimate the free energy of the structural opening reaction (ΔG_{HX}).

$$\text{NHclosed} \underset{k_{cl}}{\overset{k_{op}}{\rightleftharpoons}} \text{NHopen} \xrightarrow{k_{ch}} \text{Exchanged} \qquad \textbf{[III.3]}$$

$$\Delta G_{\mathrm{HX}} = -RT \ln K_{op} = - RT \ln (k_{\mathrm{ex}}/k_{\mathrm{ch}})$$

[III.4]

here, k_{ch} represents the chemical exchange rate of a freely accessible amide proton and depends on a variety of factors (pH, temperature, neighbouring side-chains and isotopic effects) which have been calibrated by Bai et al. (1993). The k_{ex} is the experimentally observed exchange rate for a given amide proton. The ratio k_{ex}/k_{ch} is defined as the protection factor. Assuming that the solute does not change the k_{ch} significantly, and the results from the CLEANEX-PM experiments indicate that there is less than a twofold change, the variation of the free energy values ($\delta\Delta G_{HX}$) caused by the solute can be determined using the following equation:

$$\delta\Delta G_{\mathrm{HX}} = -RT \ln K_{op} = - RT \ln (k_{\mathrm{ex}}/k_{\mathrm{ex}}^{\mathrm{solute}})$$

[III.5]

At 30°C (lowest examined temperature), the estimated free energy values for the structural opening reaction varied between 23.6 and 43.5 kJ·mol^{-1}, while at 47°C the free energy values ranged between 21.3 and 31.6 kJ·mol^{-1}. The presence of mannosylglycerate (0.25 M) caused a general increase of the ΔG_{HX} values. The values of $\delta\Delta G_{HX}$ (equation III.5) obtained at 42°C (the temperature that allowed the comparison of the largest number of ΔG_{HX} values) showed that the residues involved in α-helix 2 are significantly more stabilized than the residues in other secondary structural motifs, whereas β-strand 1 and α–helix 3, located at the N- and C-terminus, respectively, were much less stabilized by mannosylglycerate (Table III-4). The ΔG_{HX} values obtained from the 30 amide groups that were accessible at all studied temperatures were averaged to obtain a global ΔG_{HX} as a function of

temperature (Figure III-8). This shows that mannosylglycerate has a stabilizing effect over the full temperature range.

Table III-4. Added stability of SNase induced by the presence of 0.25 M mannosylglycerate as measured by the variation of proton exchange rates at 42°C. Secondary structure elements are numbered from the N-terminus to the C-terminus in the SNase sequence.

	Average ΔG_{HX} (kJ·mol^{-1}) [a]		$\delta\Delta G_{HX}$(kJ·mol^{-1})	# residues included
	no sol	MG	MG - no sol	
β-strand 1	24.83 (± 0.17)	27.81 (± 0.15)	2.98 (± 0.23)	4
β-strand 2	27.64 (± 0.08)	32.30 (± 0.14)	4.66 (± 0.16)	6
β-strand 3	29.75 (± 0.10)	34.29 (± 0.52)	4.54 (± 0.53)	6
α-helix 1	27.33 (± 0.08)	32.91 (± 0.17)	5.58 (± 0.19)	6
β-strand 4	27.06 (± 0.04)	32.42 (± 0.24)	5.36 (± 0.24)	2
β-strand 5	29.24 (± 0.07)	34.54 (± 0.27)	5.30 (± 0.28)	6
α-helix 2	29.53 (± 0.04)	35.71 (± 0.38)	6.18 (± 0.38)	7
α-helix 3	26.94 (± 0.12)	30.50 (± 0.07)	3.56 (± 0.14)	7
Global	28.04 (± 0.03)	32.70 (± 0.08)	4.66 (± 0.08)	59

[a] The values represent the average of all the residues in a particular secondary structural element that could be measured in the absence and in the presence of mannosylglycerate (MG). Associated errors were estimated from the individual uncertainties of each ΔG_{HX} value using the fundamental equation of error propagation. The abbreviation "no sol" stands for "no solute".

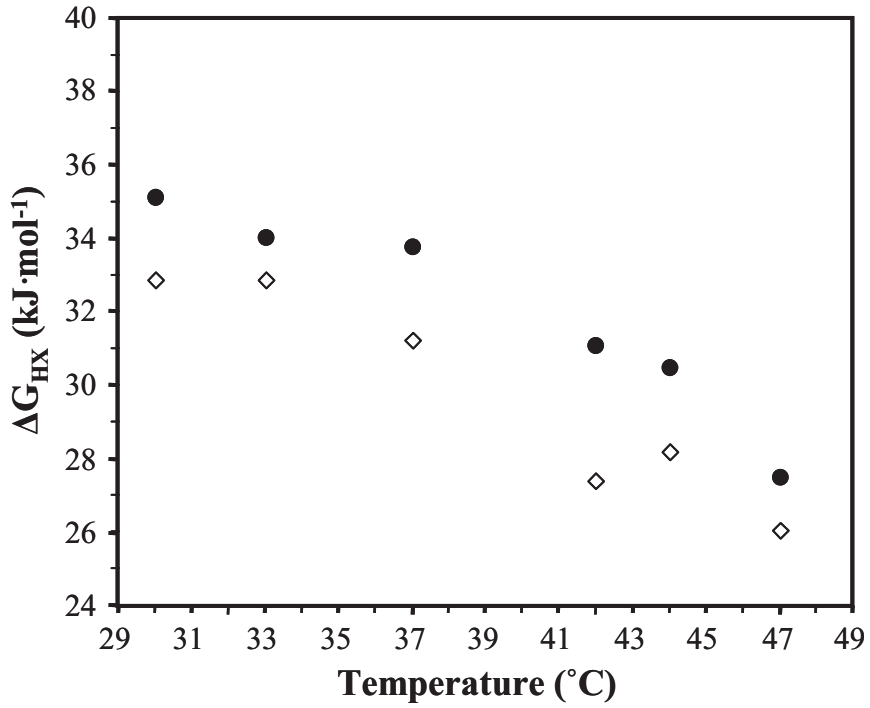


Figure III-8. Free energy of the structural opening reaction associated with H/D exchange of SNase as a function of temperature with or without mannosylglycerate. The global free energy values (ΔG_{HX}) for protein unfolding were estimated by averaging the values for the 30 amide protons with exchange rates that were accessible at all the temperatures used. The error associated with the global free energy values was lower than 0.5 kJ·mol^{-1}. Solid circles represent measurements with 0.25 M mannosylglycerate and empty diamonds those with 0.25 M KCl.

Chemical shift variation of amide protons

The chemical shift of 97 amide protons was measured as a function of the concentration of the different solutes. Mannosylglycerate caused very small chemical shift variations, but the majority of the residues showed a linear trend as a function of solute concentration, with the chemical shifts being displaced towards higher frequencies (Figure III.S-2). Fairly linear chemical shift variations were also observed with increasing amounts of the other

tested solutes. This allowed the determination of an NH solute coefficient ($\Delta\delta^{sol}$) defined as the change in chemical shift with concentration (Figure III-9). The mean coefficients were 0.22±0.04 ppm·M^{-1} for mannosylglycerate, 0.20±0.06 ppm·M^{-1} for KCl, 0.16±0.01 ppm·M^{-1} for glycerol and -0.55±0.03 ppm·M^{-1} for urea (the error values representing the standard deviation). These values show that the chemical shift variations are not correlated with stabilization of the protein.

Weak correlations (correlation coefficients $R^2 < 0.2$) were found between individual NH solute coefficients obtained with mannosylglycerate or KCl and parameters such as hydrogen-bond score (estimated by the WHATIF web interface) (Rodriguez et al. 1998), relative surface exposure (accessibility of residues as if they were mutated to an alanine) or side-chain accessible molecular surface (Figure III.S-3).

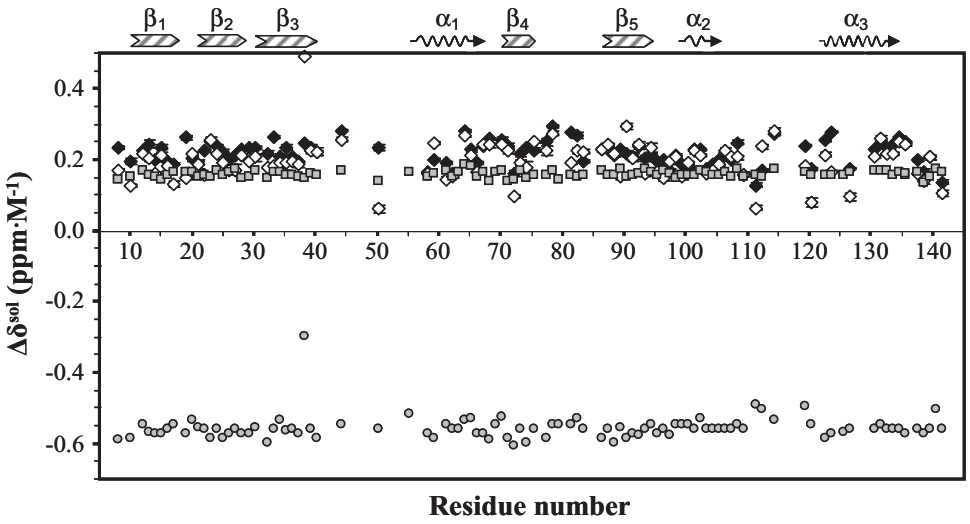


Figure III-9. Variation of the NH chemical shifts in SNase as a function of solute concentration. $\Delta\delta^{sol}$ represents the slope of the chemical shift variation for each amide proton. The values were obtained by fitting measurements made with concentrations in the range 0 to 0.25 M solute (0 to 0.6 M in the case of glycerol) at pH 5.1 (acetate buffer) and 37°C and extrapolated to 1 M. Values for KCl (empty diamonds), mannosylglycerate (black diamonds), glycerol (grey squares) and urea (grey circles) are expressed in ppm·M^{-1}.

Mannosylglycerate had little effect on the variation of NH chemical shifts with temperature (Figure III-10). The average NH temperature coefficient was found to be -3.2 ppb·K^{-1} in the presence and in the absence of mannosylglycerate. This supports the notion that solutes do not have any measurable impact on the structure of the native state.

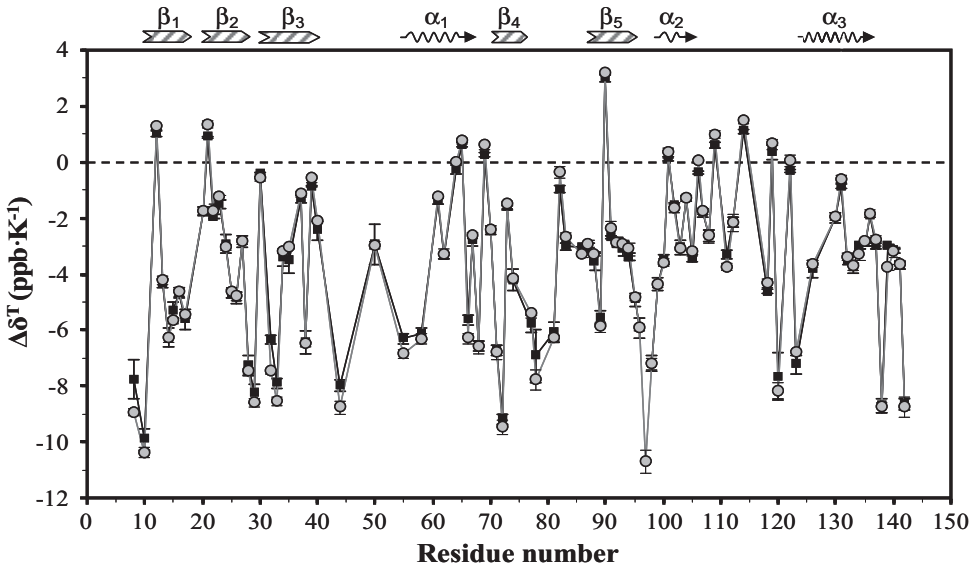


Figure III-10. Effect of mannosylglycerate on the temperature dependence of SNase amide proton chemical shifts. Measurements were made in the range 27°C to 49°C in the absence (grey circles) and presence of 0.35 M of mannosylglycerate (black squares).

DISCUSSION

Understanding the mechanisms for stabilization of proteins by compatible solutes is of great importance from a fundamental as well as an applied point of view (Cohen and Kelly 2003; Davis-Searles et al. 2001). It is generally accepted that solutes do not affect the structure of proteins in a significant way, but alterations in the pattern of protein dynamics have been reported (Doan-Nguyen and Loria 2007; Lamosa et al. 2003; Wang et al. 1995; Foord and Leatherbarrow 1998). However, the relationship between flexibility and stability remains debatable (LeMaster et al. 2005; Kamerzell and Middaugh 2008). Studying the influence of solute concentration on protein stability has a number of advantages over studies of pairs of systems such as proteins from mesophilic and hyperthermophilic organisms, point mutants, enzymes with

and without bound substrates, or oxidised and reduced centres (Krishnamurthy et al. 2009; Butterwick et al. 2004; Alexandrescu et al. 1996). In each of those cases it is difficult to separate the effects of the local differences from effects that correlate directly with protein stability. The present work allowed the dynamic properties of a single protein to be measured as its stability was varied. To ascertain whether the stabilization rendered by solutes is associated with changes in internal mobility we performed a thorough characterisation of the effect of mannosylglycerate on the internal motions of a model protein, using different NMR experiments to access multiple times scales.

A strong correlation was observed between the stability of SNase in each of the conditions studied and the respective generalised order parameters. In particular, a linear correlation was found between S^2 values and the melting temperature at increasing concentrations of mannosylglycerate (Figure III-4). Curiously, the addition of mannosylglycerate (0.35 M at 37°C) resulted in a restriction of the SNase fast motions comparable to that brought about by a 5°C reduction in the working temperature. This correlation is further strengthened by the observation that an increase in the experimental temperature, or the addition of urea, resulted in a decrease of the generalized order parameters. Interestingly, the effect of mannosylglycerate appears to be greater on residues involved in α-helices and β-strands than in less structured regions of the protein, such as loops, turns and termini, suggesting a strengthening of intramolecular interactions in the protein.

These results suggest a link between induced protein stabilization and backbone rigidification as the effect of mannosylglycerate on protein dynamics in the ps-ns time scale clearly correlates with extra stability conferred by this solute. However, this does not mean that a more stable protein will necessarily be more rigid since differences in amino acid composition may

confer added stability through other mechanisms (Petsko 2001). Indeed, comparison of homologous proteins from thermophilic and mesophilic origins has shown a marked reduction in ps-ns time scale motions in the more stable proteins in some cases but not in others (Horne et al. 2007; Krishnamurthy et al. 2009; Butterwick et al. 2004; LeMaster et al. 2005).

The CLEANEX-PM experiment is used to quantify water-NH exchange in the millisecond range (Hwang et al. 1998). The presence of mannosylglycerate decreased the values of all the exchange rates measurable by this technique by similar factors (a decrease of 1.5 to 2.5 fold). However, the viscosity of the solution appears to have a comparable effect, so it is difficult to draw conclusions about its relevance to the stabilization of SNase. As the residues in question are poorly shielded from solvent access, these small effects might be caused by the influence of the water structure in the hydration shell of the protein, *i.e.*, a change in k_{ch}, the exchange rate for unprotected protons. A study of *Pyrococcus furiosus* rubredoxin at high pH (Hernandez et al. 2000) found amide proton exchange occurring throughout the structure on the ms timescale, despite the high stability of the native protein. Similarly, the addition of 1 M sucrose to ribonuclease A increased the stability by 5°C (Liu and Sturtevant 1996) but had little effect on fast and intermediate exchange (Wang et al. 1995). Thus it seems that proton exchange on the millisecond time scale is not related to protein stability.

The most slowly exchanging amide protons are only expected to exchange with deuterium through motions that lead to partial protein unfolding. These motions are expected to be much slower and larger-scale than those characterized by the S^2 parameter. The large reduction induced by mannosylglycerate in the exchange rates of the most protected amide protons implies that the fraction of time that the protein remains in the partially unfolded state has also decreased; this means that the protein mobility has

decreased in the sense that the time-averaged structure is tighter. However, the rate constants for closing and opening are not determined (equation III-3). Exchange rates may also be affected by changes in k_{ch} but the effects observed using CLEANEX-PM place an upper limit on this and show that it is not a major factor.

Several authors defend the existence of intermediate states in the folding pathway of SNase, proposing more or less complex models (Bedard et al. 2008a; Jacobs and Fox 1994). However, the exchange reaction for each individual proton is always described by a two state model where the proton can only be in the closed or open state, with a constant k_{ch}, independent of the conformational state of the rest of the protein. In this sense, the closed state for a specific proton is a native like environment for that proton, although it may not be the native state of the entire protein. Whatever the detailed mechanism of SNase folding, the exchange rates (accessible by the method employed here) that are slowed due to the presence of mannosylglycerate imply a shift in the equilibrium towards the closed state. Hence, if a generalised reduction of the amide proton exchange rates is observed then it remains true that the time-averaged structure of the protein is more compact.

In the case of SNase, the exchange rates of the very slowly exchanging amide protons located at the center of secondary structural motifs appear to be more affected by mannosylglycerate than those located at the periphery (Figure III-11). This points to a cumulative effect of the solute on the exchange rates of the amide protons located within a structural motif or within the same sub-local unfolding unit – the foldons (Englander et al. 1997; Bedard et al. 2008b). At 42°C, half of the measured hydrogen exchange rates (i. e., twenty nine), were from residues involved in these foldon substructures. The presence of mannosylglycerate resulted in a reduction of the hydrogen

exchange rates of more than seven fold for 24 residues within foldons, whereas only four (out of twenty eight) of the residues that were not within foldons experienced similar variations of their amide proton exchange rates. This suggests that mannosylglycerate affects primarily the concerted, wider motions of the protein backbone in the latter stages of unfolding rather than more local and non-concerted motions. Furthermore, glycerol, which was used as a control for the effect of viscosity, had a relatively small influence on the exchange rates.

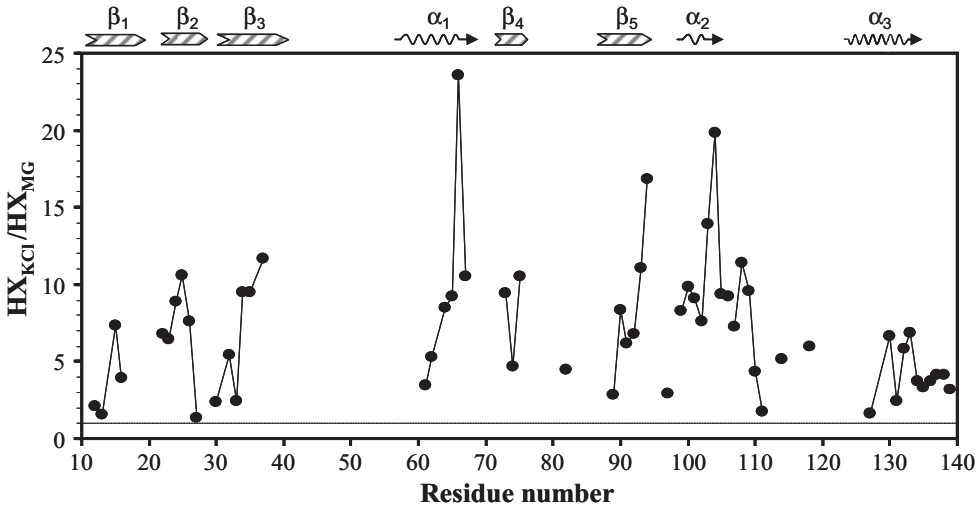


Figure III-11. Ratio of the amide proton exchange rates measured in the presence of KCl and mannosylglycerate at 42°C. The plot represents the number of times that the exchange rates of the amide protons in the presence of KCl (0.25 M) were changed by the presence of mannosylglycerate (0.25 M). Values above the horizontal dashed line represent a decrease in the exchange rates measured in the presence of KCl. The solid line is to help guide the eye.

The average values of $\delta\Delta G_{HX}$ calculated for each secondary structural motif further support this concept (Table III-4); the smallest variations of the free energies were found in β-strand 1 and α-helix 3, which are closest to the N-

and C- termini, respectively. Structural elements further away from the termini are involved in a larger number of intramolecular interactions and exhibited larger variations of the free energy values upon solute addition. In summary, mannosylglycerate restrains protein motions at this slow time scale, especially of residues in the core of the protein, as observed by other authors with stabilizing osmolytes such as sucrose, glycine, or glycerol (Wang et al. 1995; Calhoun and Englander 1985; Foord and Leatherbarrow 1998).

Mannosylglycerate causes only small changes in NH chemical shifts, which is consistent with the view that the solute does not induce appreciable modifications of the protein structure. Nevertheless, the chemical shift variations showed a linear behaviour with respect to solute concentration. A weak correlation was found with the protein relative surface exposure and side-chain accessible molecular surface (AMS) but not with backbone AMS, suggesting that the mannosylglycerate induced chemical shift variations occur via solute/side-chain interactions. However, the evidence for mannosylglycerate binding is rather weak.

CONCLUDING REMARKS

Mannosylglycerate was shown to induce restriction of the protein motions on all the time scales that were studied, but the effect on proton exchange on the millisecond time scale was weak. A strong correlation was established between restriction of the high frequency motions and the increased stability of SNase as a function of mannosylglycerate concentration, and this was not a consequence of an increase of the ionic strength or the viscosity of the solution. It should be stressed that this correlation does not imply an obligatory connection between the intrinsic stability and rigidity of individual proteins. Actually, an increasing number of experimental data have challenged

the intuitive association between highly stable proteins and high rigidity (LeMaster et al. 2005; Kamerzell and Middaugh 2008). Further studies on the effect of stabilizing solutes on protein dynamics will be necessary to determine how general the established link between protein rigidification and stabilization is.

ACKNOWLEDGEMENTS AND WORK CONTRIBUTIONS

This work was supported by Fundação para a Ciência e a Tecnologia (FCT), Portugal, Projects POCI/BIA-PRO/57263/2004 and PTDC/BIO/70806/2006. The authors acknowledge Dr. Tiago Faria, ITQB, for valuable help with the statistics analysis and protein production. Tiago M. Pais acknowledges FCT for grant support (SFRH/BD/42210/2007). The NMR spectrometers are part of the National NMR Network and were acquired with funds from FCT and FEDER.

SUPPLEMENTARY MATERIAL

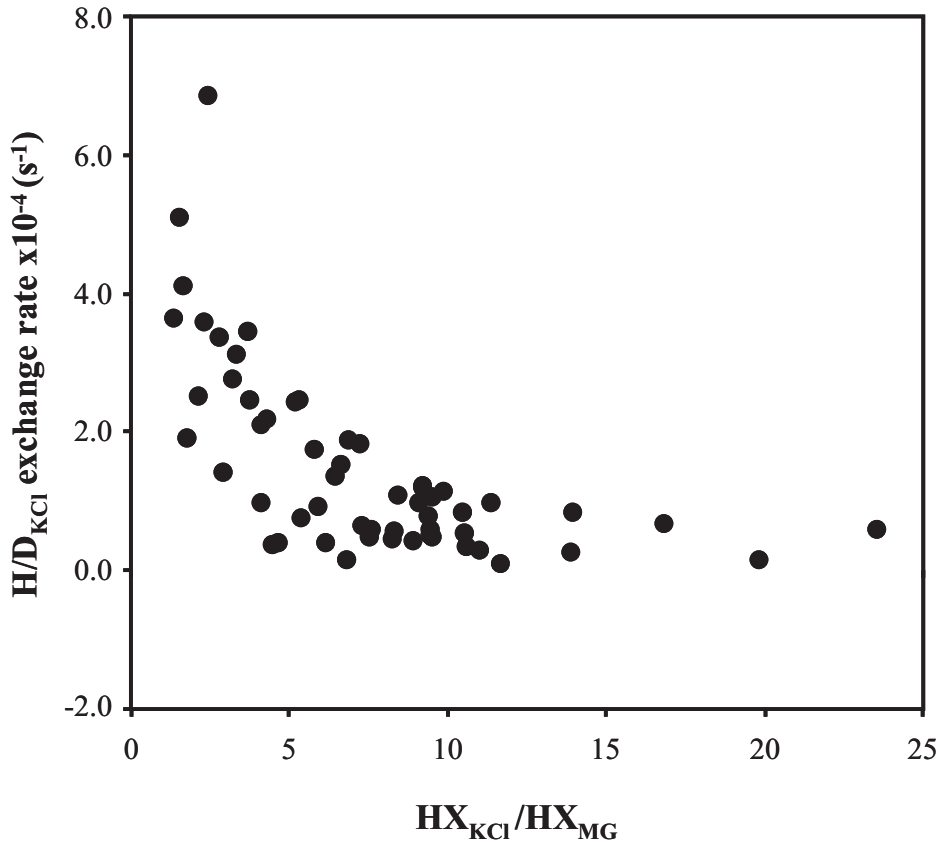


Figure III.S-1. Inverse correlation between intrinsic exchange rates of amide protons and their modulation by mannosylglycerate. The plot shows the correlation between intrinsic exchange rates at 42°C and the factor by which these rates were changed in the presence of mannosylglycerate (0.25 M). The abscissa represents the ratio between the exchange rate of each amide in the presence and absence of mannosylglycerate (0.25 M), while the ordinate represents the intrinsic exchange rates of each amide measured in the presence of KCl (0.25 M).

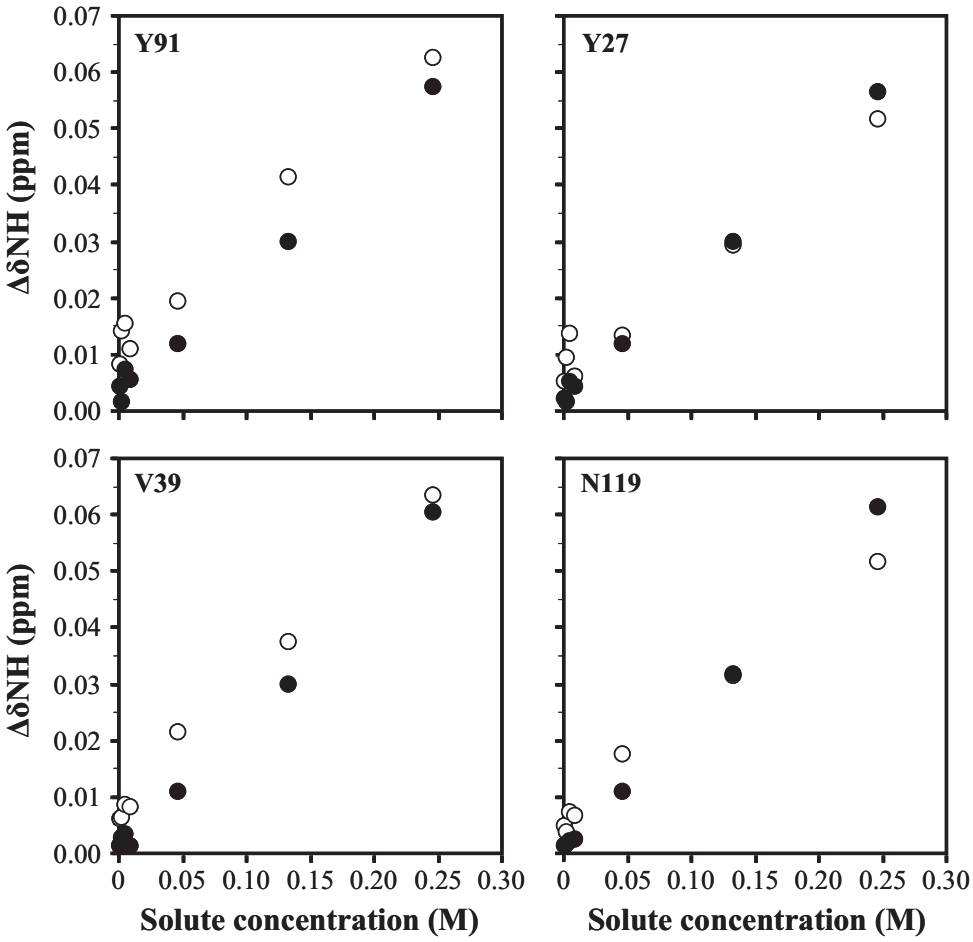


Figure III.S-2. Chemical shifts of amide protons as a function of solute concentration. The plots represent the effect of increasing amounts of mannosylglycerate (solid circles) or KCl (open circles) on the chemical shifts of four selected amide protons. Chemical shifts were referenced to dioxane in a capillary tube.

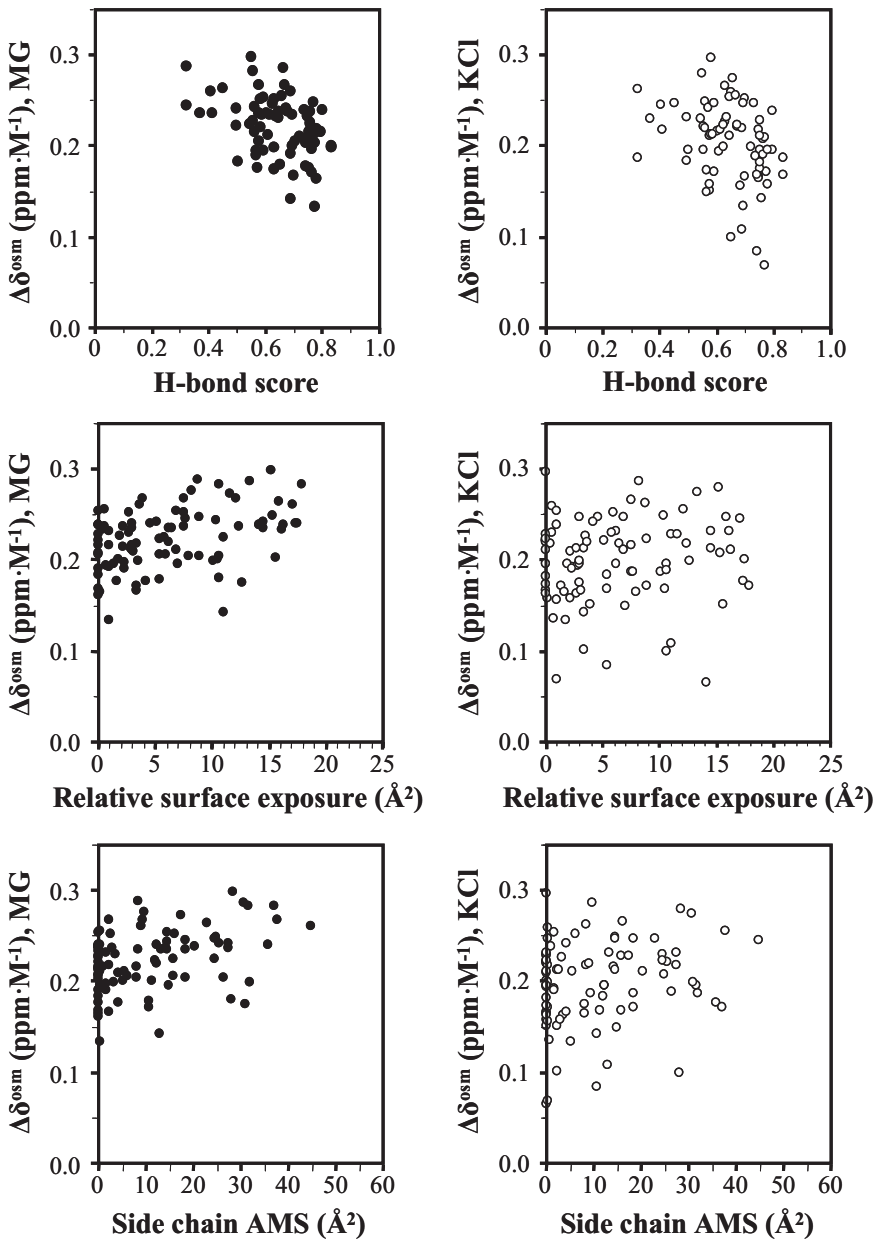


Figure III.S-3. Dependence of amide proton chemical shifts on solute concentration and its correlation with various structural parameters. The effect of mannosyglycerate (filled circles) or KCl (empty circles) is compared with hydrogen bond score (Rodriguez et al. 1998) (upper panel), relative surface exposure (middle panel) and side-chain accessible molecular surface (lower panel)

Chapter IV

Mannosylglycerate stabilizes SNase with specific restriction of slow β-sheet motions

CONTENTS

ABSTRACT

Mannosylglycerate is a compatible solute typical of thermophilic microorganisms that has a remarkable ability to protect proteins from thermal denaturation. To understand how mannosylglycerate confers protection, we have been studying its influence on the internal motions of a hyperstable staphylococcal nuclease (SNase). Previously, we found a correlation between the magnitude of protein stabilization and the restriction of backbone motions. We now report the effect of mannosylglycerate on the fast motions of side-chains and on the slower unfolding motions of the protein. Side-chain motions

were assessed by $^{13}CH_3$ relaxation measurements and model-free analysis while slower unfolding motions were probed by H/D exchange measurements at increasing concentrations of urea. Side-chain motions were little affected by the presence of different concentrations of mannosylglycerate (up to 0.35 M) or even by the presence of urea (0.25 M), and show no correlation with changes in the thermal stability of SNase. Native hydrogen exchange experiments showed that, contrary to reports on other stabilizing solutes, mannosylglycerate restricts slower local motions in addition to the global motions of the protein. The intermediate states in the protein unfolding/folding pathway remained undisturbed in the presence of the solute but mannosylglycerate showed a specific effect on the local motions of β-sheet residues.

INTRODUCTION

To understand the mechanism of protein stabilization by compatible solutes we have been studying the effect of mannosylglycerate on the dynamics of a hyperstable staphylococcal nuclease variant (SNase). SNase is a calcium dependent phosphodiesterase expressed by *Staphylococcus aureus* to hydrolyse DNA and RNA to 3'-mononucleotides (Tucker et al. 1978). Mannosylglycerate is a charged compatible solute found in many hyperthermophilic bacteria which was shown to greatly enhance protein resistance to thermal denaturation (Faria et al. 2008; Ramos et al. 1997). Since the pioneering studies by Tanford and Nozaki (Nozaki and Tanford 1963; Nozaki and Tanford 1970), many researchers have made important contributions (Arakawa and Timasheff 1985; Auton and Bolen 2004; Liu and Bolen 1995; Timasheff 1992) to the field of protein stabilization by osmolytes. There is now a much better understanding of the main features, but some

areas remain little explored and some observations (Pais et al. 2005) are difficult to explain with any of the existing models.

In particular, further studies are necessary to clarify the relationship between protein dynamics and stabilization by compatible solutes. Previously, the effect of different solutes on a wide range of backbone motions was monitored by using several NMR methods. Small fluctuations of bond lengths (ps-ns time scale) were studied by spin-relaxation measurements, chemical exchange on the millisecond scale was assessed by magnetization transfer experiments, and events on the second to minute time frame were probed with proton/deuterium amide exchange experiments (Pais et al. 2009). A strong correlation was established between the sub-nanosecond backbone motions (described by the generalized order parameter) and changes in the melting temperature (T_m) of the protein induced by different solutes. The effect of mannosylglycerate was also evident on the slowest time scale, providing site-specific information about the thermodynamic stability of the protein. However, motions on the millisecond time scale were little affected by the different solutes. Thus, there was strong evidence in this case for a link between restriction of protein backbone motions and protein stabilization.

Following these results, we consider it important to address the following questions. Is the dynamic behaviour of protein side-chains on the ns-ps timescale also important for the mechanism of protein stabilization by compatible solutes? Is the effect of stabilizing solutes really specific for the slower global motions of the protein, as previously suggested (Wang et al. 1995)? And do stabilizing solutes affect the protein folding/unfolding pathway?

To study the effect of solutes on the dynamics of protein side-chains, we measured ^{13}C relaxation rates of the methyl groups of a staphylococcal nuclease variant (P117G/H124L/S128A), providing 46 probes of motion on the sub-nanosecond time scale. Although they are typically located in the close

packed interior of proteins, methyl bearing side-chains usually display significant motions in addition to the rotation of the methyl group. Analysis of ^{13}C relaxation data using the model-free formalism (Lipari and Szabo 1982a; Lipari and Szabo 1982b) is a common procedure to extract information about the motion of the carbon-carbon bond that connects the methyl group to the side-chain, and the rotation motion of the C-H vectors about the symmetry axis. The methyl groups are expected to report on the effects of different solutes on the dynamics of the protein hydrophobic core; this work will contribute to a better understanding of the response of sub-nanosecond SNase motions to the presence of stabilizing and destabilizing solutes.

The thermodynamic equilibrium of proteins implies a constant unfolding and refolding of the molecules, even under conditions that favour the folded form. This creates transient non-native states that represent a minute fraction of the protein population under normal conditions and allows the protein folding/unfolding pathway to be studied using NMR hydrogen/deuterium exchange experiments. Most backbone amide protons exchange with the solvent only when the protein molecule is in one of the non-native states. In the so-called *EX2* exchange regime, proton exchange rates depend on the equilibrium constant between the native and non-native states (Hvidt and Nielsen 1966). The motions involved in visiting these states may be local motions that expose little new surface area, or nearly global motions that expose large patches of normally buried peptide segments (Hvidt and Nielsen 1966). Stabilizing solutes reduce the exchange rates of amide protons, which implies a shift in the equilibrium of the protein population towards the native state (Wang et al. 1995; Pais et al. 2009; Lamosa et al. 2003). The observation of a greater effect on the amide protons that exchange more slowly even in the absence of solutes has been explained as a consequence of solutes opposing large increases in protein surface area while the slower

exchanging protons are expected to exchange only through large scale unfolding motions. However, it is difficult to establish a quantitative correlation between the effect of stabilizing solutes and exchange rates, possibly because not all slower-exchanging amide protons require large protein motions. Investigating the scale of the motions associated with amide proton exchange may clarify the matter and deal with our second query: is the effect of stabilizing solutes really specific for the global motions of the protein?

The dependence of individual hydrogen exchange rates on denaturant (urea) concentration can be used to differentiate between local and global backbone motions. The slope of this dependence, called the *m*-value, relates to the increase in surface exposure that occurs when the protein visits the higher energy exchange-competent state while remaining under largely native conditions (folded population around 99%) (Tanford 1970; Schellman 1987a). Values close to zero are characteristic of local motions while large *m*-values indicate that amide exchange requires nearly global motions (Bai et al. 1995; Mayo and Baldwin 1993). Englander and co-workers found that protein assembly proceeds via cooperative folding of protein segments which they called foldons (Bai et al. 1995), with hydrogen exchange measurements used to show which amides are involved in each foldon. Experiments of this type in the absence and presence of mannosylglycerate should address our last question regarding the possible effect of stabilizing solutes on the cooperative unfolding motions of the protein.

By addressing these questions, this work shows stabilization of SNase by mannosylglycerate has little effect on side-chain dynamics and restricts local fluctuations of amide protons as well as global opening motions. However, there is an intriguing insight into the mechanism of stabilization in the observation of specific effects of mannosylglycerate on the local motions of the β-sheets in SNase.

MATERIALS AND METHODS

Preparation of Protein Samples

The protein used for these studies is the hyperstable PHS variant of SNase which differs from the wild type amino acid sequence in three positions: P117G, H124L, S128A (Calhoun and Englander 1985). The expression and purification protocols for SNase samples were as described elsewhere (Pais et al. 2009; Calhoun and Englander 1985), except for the medium composition which differs in the nitrogen and carbon sources. For the uniformly $^{13}C/^{15}N$ labelled SNase sample, [$^{13}C_6$]-glucose (4 g/L) and $^{15}NH_4Cl$ (2 g/L) were used as sole carbon and nitrogen sources. Fractionally ^{13}C labelled SNase was prepared using a mixture of 15% [2-^{13}C]acetate, 15% [1-^{13}C]acetate and 70% [1,2-^{12}C]acetate at a final concentration of 4 g/L, as sole carbon source, and $^{14}NH_4Cl$ (2 g/L) as nitrogen source. This labelling strategy results in a protein molecule where carbon sites have a 15% probability of being labelled with ^{13}C. Finally, uniformly ^{15}N labelled SNase samples were obtained using $^{15}NH_4Cl$ (2 g/L) as sole nitrogen source while unlabeled SNase was obtained using LB rich medium. The concentration of purified protein stocks was estimated by the solution absorbance at 280 nm and the extinction coefficient of 0.93 ($cm^{-1}\cdot mg^{-1}\cdot ml$) (Faria et al. 2004). The protein purity was assessed by SDS-PAGE electrophoresis with Coomassie staining. Purified protein stocks were flash-frozen with liquid nitrogen in small drops as this reduces protein aggregation, and stored at -80°C.

Circular Dichroism Spectroscopy

Circular dichroism (CD) experiments were performed at 35°C on a Jasco J-815 spectropolarimeter equipped with a Peltier thermostatted cell. Spectra of 15 μM SNase samples (acetate buffer, 60 mM, pH 5.5) were recorded in the

range of 200-250 nm using a 0.1 mm path length cell and represent the average of 7 scans. Chemical unfolding of SNase was studied by monitoring the CD signal intensity at 222 nm in the presence of several urea concentrations (0 – 6 M). Non-linear least square fitting of the data was carried using the Santoro-Bolen equation (Santoro and Bolen 1988).

NMR Spectroscopy

All NMR experiments were performed at the ITQB magnetic resonance center, CERMAX. Spectra for the assignment of SNase resonances and for ^{13}C relaxation measurements were collected at 25°C on a Bruker Avance*III* 800 spectrometer (Bruker, Rheinstetten, Germany) operating at 800.33 MHz, equipped with a TXI-Z H C/N/-D (5 mm) probe. Experiments for HX measurements were performed on a Bruker AVANCE*III* 500 spectrometer (Bruker, Rheinstetten, Germany), operating at 500.13 MHz, using a 5 mm QXI inverse detection probe-head with pulsed-field gradients along the Z axis.

Data were processed with Topspin 2.1 (Bruker Biospin) and analyzed with CARA version 1.8.4. Backbone sequential assignments were obtained from the analysis of ^{1}H-^{15}N HSQC and triple resonance HNCA, HNCO, HN(CO)CA, CBCA(CO)NH, and HNCACB experiments. Side-chain resonances were assigned with the combined use of ^{15}N HSQC-TOCSY, ^{15}N HSQC-NOESY, HBHA(CO)NH, (H)CCH-TOCSY, and ^{1}H-^{13}C HSQC-NOESY spectra. In addition for the aromatic side-chains we used the ^{1}H-^{13}C HSQC-NOESY and (HB)CB(CGCD)HD spectra. The presence of H-bonds was deduced from the H-X data and added according to the secondary structure and the structures of the first structure calculations which were performed in their absence.

Structure Determination

Initial models of the structure were calculated with the program UNIO'10 (Herrmann et al. 2002a; Herrmann et al. 2002b) using the ^{1}H-^{1}H NOESY, 3D

^{1}H-^{15}N HSQC-NOESY, 3D ^{1}H-^{13}C HSQC-NOESY spectra and chemical shift lists tailored to each spectrum. For each calculation cycle, 100 structures were produced using CYANA2.1 protocols (Güntert et al. 1997), with 10000 simulated annealing steps. In this procedure, the residue ranges of 6-42 and 56-145 were used for the calculation of the RMSD. At the end of the seven iterative calculation cycles, bundles of 20 structures were generated, together with upper limit constraint lists. These constraint lists were then used for de novo calculations of structures in explicit water using the RECOORD protocols (Nederveen et al. 2005). First, 80 structures were calculated using the included CNS protocols (Nederveen et al. 2005). The best twenty structures were then used for explicit water calculations and from those, twenty were selected based on the smallest number of NOE violations and the ten with lowest calculated energy were used to form the final NMR bundle.

The quality of the calculated structures was evaluated using the suit of programs iCING (https://nmr.cmbi.ru.nl/icing/iCing.html).

^{13}C relaxation measurements

^{13}C relaxation rates were measured for the methyl groups of 1 mM SNase samples fractionally labelled with ^{13}C (60 mM acetate-d_4 buffer, pH 5.1), at 37°C in the absence and presence of the following compounds: mannosylglycerate (0.15, 0.25, and 0.35 M), KCl (0.25 M), glycerol (0.60 M), and urea (0.25 M). The concentrations of KCl and glycerol are chosen to match the ionic strength and viscosity of 0.25 M mannosylglycerate, respectively.

The pulse sequences used to measure ^{13}C longitudinal (T_1), and transverse (T_2) relaxation times and steady-state heteronuclear {^{1}H-^{13}C}NOE (h-NOE) were based on published schemes (Nicholson et al. 1992) with the following modifications: gradient pulses (sine shape) were included and the last carbon

pulse was a 180º adiabatic shaped pulse (Crp60,0.5,20.1) of 500 μs; spectra for T_1 and T_2 relaxation measurements were acquired as pseudo 3D data while spectra for h-NOE determination were acquired in an interleaved fashion; the T_2 pulse sequence includes a CPMG heat compensation block so that all time points are acquired in similar conditions.

Longitudinal relaxation rates (R_1) were measured using relaxation periods of 10, 50, 100, 250, 350, 600, 800, and 1200 ms while transverse relaxation rates (R_2) were obtained using relaxation periods of 14, 28, 42, 56, 70, 83, 97, and 111 ms. Peak volumes were determined using the integration mode of TopSpin 2.1 software (Bruker) and their time dependence fitted to a mono exponential (for R_1) or biexponential function (R_2) by non-linear least squares. The transverse ^{13}C relaxation of an isolated methyl group rapidly rotating about its threefold axis and attached to a macromolecule is better characterized by a biexponential decay of the form $y = 0.5A[\exp(-T/T_{2f}) + \exp(-T/T_{2s})]$ due to the inner and outer ^{13}C multiplet components (Nicholson et al. 1992). In this equation, T_{2f} and T_{2s} represent the fast and slow components of the transverse relaxation rate. The value used for the subsequent model-free analysis is the average of the relaxation rates, $T_2 = [0.5(T_{2s} + T_{2f})^{-1}]^{-1}$. Error estimates were obtained from the standard deviation of the fitted coefficients. Experiments for h-NOE measurement were performed in duplicate and the respective value determined from the ratio of the peak volume of each resonance in the presence and absence of ^1H saturation (V_{sat}/V_{nonsat}). The NOE was allowed to build for 5 seconds. For the control experiment (no NOE) a minimum relaxation delay of 7 seconds was used.

Model-free analysis of ^{13}C relaxation data

Relaxation data was analyzed with the RELAX software package (d'Auvergne and Gooley 2008a; d'Auvergne and Gooley 2008b) and the Lipari-Szabo formalism (Lipari and Szabo 1982b; Lipari and Szabo 1982a). The dynamics of ^{13}C labelled methyl groups were characterized by one of 5 models: 1) S^2; 2) S^2 and τ_e; 3) S^2 and R_{ex}; 4) S^2, τ_e, and R_{ex}; 5) S^2_s, S^2_f, and τ_s. For each methyl group, the best model was selected based on the lowest Akaike's information criteria (AIC). The fifth and most complex model considers two contributions, a fast motion (S^2_f) corresponding to methyl group rotation about its symmetry axis (the carbon-carbon bond connecting the methyl group to the side-chain), and a slower motion (S^2_s) with a correlation time, τ_s, corresponding to the reorientation of the symmetry axis (Woessner 1962; Palmer, III et al. 1993). The simpler models consider a fast motion described by the generalised order parameter, S^2, and assess the presence of chemical exchange (R_{ex}) and effective correlation time (τ_e).

Hydrogen exchange experiments

Samples used in hydrogen exchange experiments contained 1 mM of ^{15}N labelled SNase, 0.15, 0.24, 0.33 M mannosylglycerate or no mannosylglycerate, and urea at several concentrations. The reaction of proton/deuterium exchange was initiated by dissolving lyophilized protein samples in deuterated solutions (acetate-d_4 buffer, 60 mM, pD 5.2) with increasing concentrations of urea. The sample was thermally equilibrated at 35°C for 10 min in the spectrometer before collecting sequential {^1H-^{15}N}-HSQC spectra. The first spectrum started less than 15 minutes after the exchange reaction was initiated. Spectra were acquired with 1024 complex data points, 136 time increments, and 4 scans per increment and each took less than 15 min.

Cross-peak volumes of the NH resonances present in the HSQC spectra were measured with TopSpin 2.1 software (Bruker) and fitted to a mono-exponential decay function using MatLab and the Solver package of Excel, providing the exchange rate for each detected amide proton resonance.

RESULTS

Assignment and solution structure of the PHS variant of SNase

We have identified 143 out of 149 possible amide resonances in the {^{1}H-^{15}N}-HSQC spectrum of SNase (Figure IV.S-1). Residues A1, T2, K6, K53, Y54 and E57 could not be observed in the ^{1}H-^{15}N HSQC, however with the exception of A1 and K53, the side-chains of these residues have been identified in all the ^{13}C HSQC based spectra. All the missing amide resonances originate from amino-acids located in regions of high flexibility, which is to be expected since proton exchange is likely to be relatively slow at the low pH (5.2) at which the NMR data were collected.

Overall 84% of the possible resonances were assigned. More specifically, 93% of backbone assignments, 90% of the aliphatic ^{1}H and ^{13}C and 45% of all aromatic protons were identified. A small number of groups with resonances far removed from the average values seen in the BMRB database are present. These are K9 (HE2/HE3), D77, and M98. The uncommon values of HE of K9 are likely caused by its proximity to the aromatic ring of Y92.

The solution structure of the PHS variant of SNase is very similar to previously published crystal structures. The RMSD between the average NMR structure and the crystal structure (PDB entry 1EY8) is 0.96 Å. The main differences are in the loop spanning residues 42-52 (Figure IV.S-2). This region exhibits the lowest order parameters, indicating that the loop is flexible (Pais et al. 2009). As can be seen in Figure IV-1, this region is not well

defined in the NMR structure, since very few meaningful NOEs can be extracted for that region. Further differences are observed is in the 25-33 range, encompassing the loop connecting sheets 2 and 3. This part of the structure is well defined and the reason for this structural discrepancy is not immediately evident.

The chemical shift assignments have been deposited in the BioMagResBank (http://www.bmrb.wisc.edu) with the accession number 18013, and the structures were deposited at RCSB (PDB ID: 2lkv).

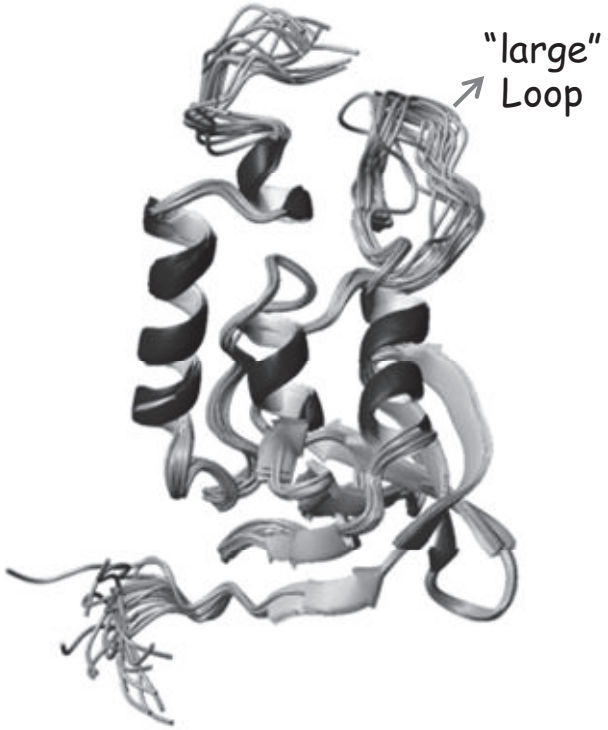


Figure IV-1. Ribbon representation of the 20 best calculated structures of a SNase variant (P117G, H124L, S128A) calculated at 37°C. Figure prepared with MolMol (Koradi et al. 1996). The N-terminal is at the bottom of the picture and the C-terminal is at the top. The less-structured loop spaning resiudes 42 to 53 is indicated in the figure as "large" loop.

Relaxation data and diffusion tensor

The relaxation rates of 46 $^{13}CH_3$ resonances from 30 amino acid side-chains were measured. Signal to noise ratios were generally above 30 for the peaks in the first time point spectra; slightly lower sensitivities were obtained for the NOE experiment without proton saturation. Longitudinal relaxation decay curves were well fit with single exponentials. Most of the transverse relaxation decay curves were best represented by a biexponential function (Nicholson et al. 1992) as assessed by *F*-test analysis. Uncertainties in the optimized parameters of R_1 and R_2 are generally below 10% while the standard deviation of duplicate NOE experiments was, generally, no greater than 5%.

Model-free analysis of ^{13}C relaxation data made use of the diffusion tensor parameters obtained from ^{15}N relaxation measurements performed previously under the same experimental conditions (Pais et al. 2009). Specific diffusion tensor values were used for each experimental condition: no solute, KCl, glycerol, urea, and mannosylglycerate at 0.15, 0.25 and 0.35 M; fully anisotropic diffusion was found to be a better fit in each case. Since ^{15}N relaxation measurements provided a larger data set, it is expected that the diffusion tensor values are more precise than could be obtained from the ^{13}C relaxation data. The solution structure of the SNase variant determined by NMR spectroscopy was used as the structural model for the model-free calculations, with coordinates adjusted to the principal axes of the diffusion tensor using the program Tensor2 (Cordier et al. 1998; Dosset et al. 2000). Additionally, the coordinates of the methyl protons were averaged along the symmetry axis (the axis described by the C-C bond connecting the methyl group to the side-chain).

Sequential fitting of the ^{13}C relaxation data to the five model-free models was performed using the RELAX software (d'Auvergne and Gooley 2008a; d'Auvergne and Gooley 2008b), assuming an inter-nuclear ^{1}H-^{13}C distance of

1.07 Å and negligible effect of chemical shift anisotropy. The complete RELAX script is provided as supplementary material (Figure IV.S-3).

Internal mobility

In the absence of solutes, the majority of the methyl groups (42 out of the 46) required the so-called extended model of the Lipari-Szabo formalism (S^2_s, S^2_f, and τ_s) to explain the relaxation data. For the remaining groups, the less complex models two (S^2 and R_{ex}) and four (S^2, τ_e, and R_{ex}) were sufficient to describe the data. In some cases, S^2 values were found to depend on the model-free model used to fit the relaxation data. Therefore, we consider only the methyl group datasets that were fitted with the same model in all experimental conditions: 36 methyl groups which were all fitted with model 5 (S^2_s, S^2_f, and τ_s).

The average τ_s for these methyls is below 0.15 ns and the slowest value is just under 0.25 ns, whereas τ_c values determined for this SNase variant are larger than 6 ns. Interpretation of relaxation data with the model-free formalism requires that τ_s values are at least 10 times smaller than the global tumbling time so that S^2 values are not biased by the slower τ_c.

The dynamics of side-chains, as represented by differences in the generalised order parameter of methyl groups in Figure IV-2, seems largely unaffected by the presence of stabilizing or destabilizing solutes. Most of the changes observed in the S^2 values are within the uncertainty limits estimated by Monte Carlo simulations. The significant variation observed for Ile72δ and Leu124δ, coincides with methyl groups displaying S^2 values above the theoretical limit of 0.111 for perfect tetrahedral geometry of the methyl group. More modest S^2 variations occur for Ile18γ, Val66γ^2, Ile92δ and Val104γ^2 methyl groups but with no apparent correlation with the concentration of mannosylglycerate (not shown). In the case of the two valines, similar S^2

variations are observed in the presence of glycerol (0.60 M) and KCl (0.25 M), two non-stabilizing compounds used, respectively, as controls of viscosity and ionic strength. The location of these side-chains in the protein structure is not associated with a particular region of the protein or to the degree of solvent exposure.

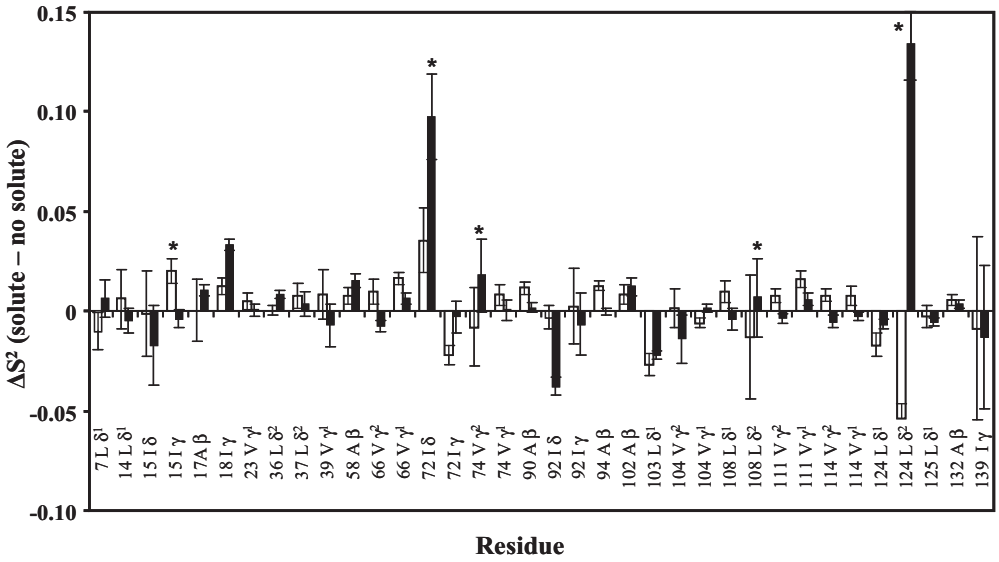


Figure IV-2. Changes in the generalised order parameters (ΔS^2) of the methyl groups induced by the presence of mannosylglycerate (0.35 M) and urea (0.25 M). ΔS^2 is obtained by subtracting the S^2 values of the SNase in the presence of mannosylglycerate (full bars) or urea (empty bars) from the values determined in the absence of solutes. The asterisks above some of the bars identify methyl groups with S^2 values above the theoretical limit of 0.111.

The average S^2 value of the side-chains clearly does not respond to changes in protein stability in the same way as the backbone dynamics determined in earlier work (Pais et al. 2009) (Figure IV-3).

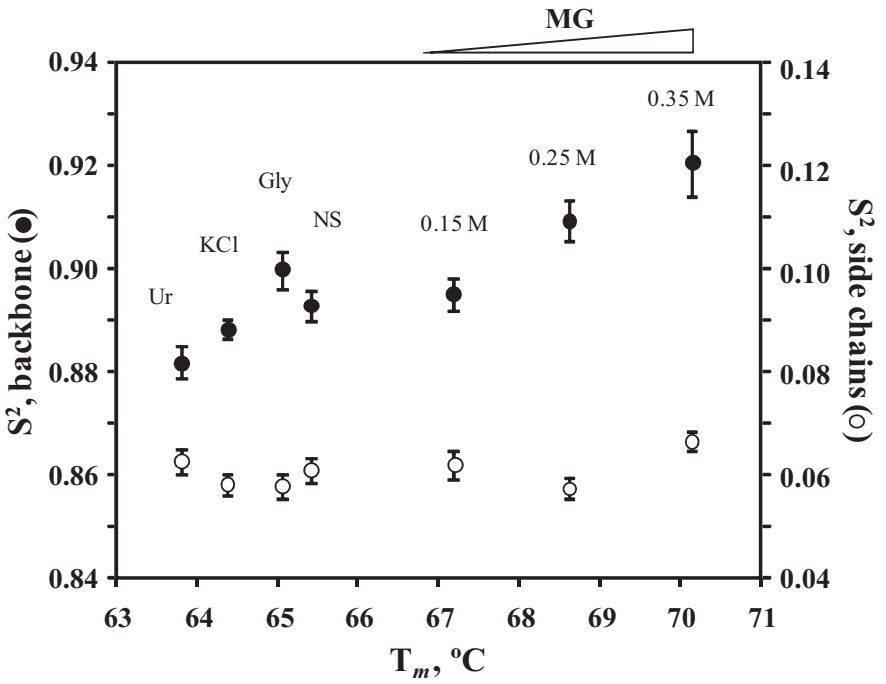


Figure IV-3. Comparison of the average S^2 values of the backbone amides and methyl groups as a function of the SNase melting temperature. The left vertical axis shows the S^2 values of backbone (full circles) and the right axis shows the S^2 values of the side-chains bearing methyl groups (empty circles). The melting temperature (T_m) of SNase is shown for the protein in the presence of a given solute: no solute (NS), KCl 0.25 M (KCl), glycerol 0.60 M (Gly), urea 0.25 M (Ur), and mannosylglycerate (MG) at 0.15 M, 0.25 M, and 0.35 M. The error bars represent the error of the average calculated from the individual S^2 uncertainties.

Analysis of Hydrogen exchange

The exchange reaction of an amide proton occurs only in open conformations (NHopen) in which the proton is accessible to the solvent, as in equation IV.2 (Hvidt and Nielsen 1966; Linderstrøm-Lang 1955).

$$\mathrm{NH(closed)} \underset{k_{cl}}{\overset{k_{op}}{\rightleftharpoons}} \mathrm{NH(open)} \xrightarrow{k_{ch}} \mathrm{Exchanged} \qquad \textbf{[IV.2]}$$

In the steady state, the experimentally observed exchange rate (k_{ex}) for any given amide proton is described by equation IV.3.

$$k_{ex} = (k_{op} \times k_{ch})/(k_{op} + k_{cl} + k_{ch}) \qquad \textbf{[IV.3]}$$

In the bimolecular regime (*EX2*) typical of stable folded proteins, the closed state (NHclosed) is highly favoured. The closing rate (k_{cl}) is much larger than the opening rate (k_{op}) and also much larger than the exchange rate of a freely accessible amide proton (k_{ch}). With these approximations, an equilibrium constant of opening can be defined and the free energy of the structural opening reaction (ΔG_{HX}) can be derived from the observed exchange rate, (k_{ex}) (equations IV.4 and IV.5).

$$k_{ex} = (k_{op} / k_{cl}) \times k_{ch} = K_{op} \times k_{ch} \qquad \textbf{[IV.4]}$$

$$\Delta G_{HX} = - RT \ln K_{op} = - RT \ln (k_{ex}/k_{ch}) \qquad \textbf{[IV.5]}$$

The values for k_{ch} can be calculated as described in Bai et al. (Bai et al. 1993) so that the free energies may be obtained from the measured exchange rates.

Amide exchange may result from local fluctuations, involving single hydrogen-bond breaks, or from global opening motions of the protein. These modes make different contributions to the free energy of exchange (Bai et al. 1995) which may be distinguished by studying the dependence of ΔG_{HX} on the concentration of denaturant. Local motions expose little new surface area of the protein to the solvent, making them relatively insensitive to the denaturant concentration, whereas global opening motions increase significantly the protein surface area and expose additional denaturant binding sites. As a result, global motions become dominant at higher concentrations of denaturant.

Two possible models were considered to describe the dependence of ΔG_{HX} with added denaturant (urea). In these, exchange occurs either through global (g) motions alone as in equation IV.6, or both local (lc) and global motions contribute, as in equation IV.7 (Mayo and Baldwin 1993; Bai et al. 1995).

$$\Delta G_{HX} = - RT \ln \{\exp((m_2[urea] - \Delta G_{HX}(g))/RT)\} \quad \textbf{[IV.6]}$$

$$\Delta G_{HX} = - RT \ln \{\exp((m_1[urea] - \Delta G_{HX}(lc))/RT) + \exp((m_2[urea] - \Delta G_{HX}(g))/RT)\} \quad \textbf{[IV.7]}$$

Here, $\Delta G_{HX}(g)$ and $\Delta G_{HX}(lc)$ are the free energies at zero denaturant concentration of the global and local opening reactions, respectively, and m_i is the dependence of the ΔG_{HX} on the denaturant concentration (i=1 for local motions and i=2 for global motions). These m-values correlate with the additional denaturant binding sites of the protein in the transient open state (Tanford 1970; Schellman 1987a; Pace et al. 1996). An F-test ($\alpha = 0.10$) was used to establish whether the use of the more complex model to describe the experimental data was statistically significant, *i.e.*, if more than one opening mode contributes to the exchange reaction.

Dependence of Hydrogen Exchange rates on urea concentration

The signal to noise ratio of the resonances in the first HSQC spectra was 20 to 30, allowing accurate determination of peak volumes through most of the decay time and good precision in the fitted exchange rate values. The time decay of the amide resonances observed at 35°C was generally well described

by a mono-exponential function and the profile was not changed by the presence of urea or mannosylglycerate.

The exchange rates of 68 backbone amide protons were determined, accounting for about 50% of the total. The presence of mannosylglycerate made the decay of some amide resonances too slow to be accurately determined within the experimental time (ca 60h) but these became measurable as the concentration of urea increased.

The urea unfolding curve of SNase determined by circular dichroism (Figure IV-4) shows that 99% of the protein population is still folded at 2 M urea, the highest concentration used in the exchange rate measurements without mannosylglycerate, indicating that the *EX2* exchange regime should still hold under these conditions. Therefore, equation IV.5 was used to extract ΔG_{HX} values for the amide protons.

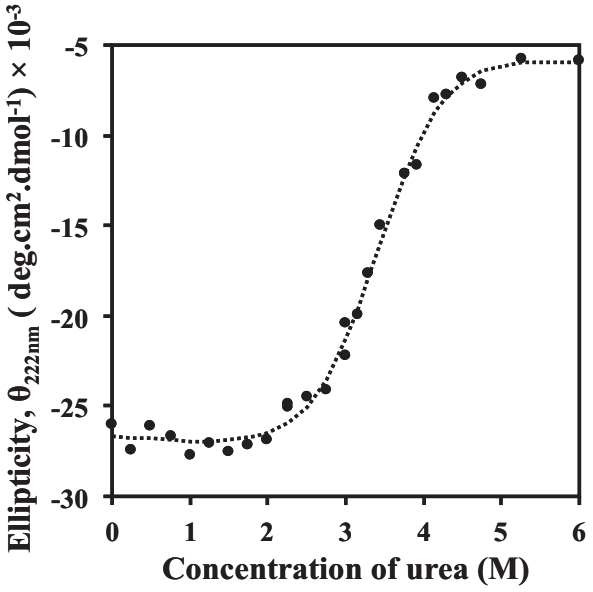


Figure IV-4. Urea unfolding of the hyperstable PHS SNase variant measured by circular dichroism at 222 nm. Protein concentration was 15 µM in acetate buffer (60 mM), at pH 5.5 and 35°C. The data was fitted using the Santoro-Bolen equation (Santoro and Bolen 1988).

The variation of ΔG_{HX} values with the concentration of urea shows four types of profile, as illustrated in Figure IV-5. Profile 1 is characterized by a two parameter model (equation IV.6), m_2 and ΔG_{HX}(g), while profiles 2 and 3 are characterized by a four parameter model (equation IV.7): m_1 and m_2, and ΔG_{HX}(lc) and ΔG_{HX}(g). Profile 3 differs from profile 2 because its m_1 value is essentially zero. The less common profile 4, illustrated by the nearly flat line of Arg87 in Figure IV-5, is a limiting case of profile 2 (Table IV.S-1 lists the values of the model parameters determined for each amide proton).

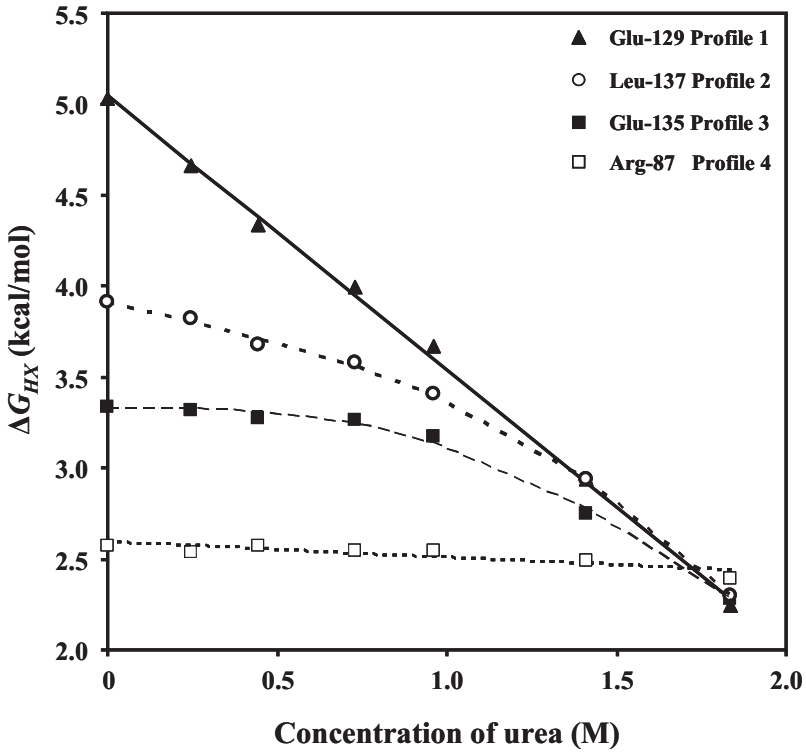


Figure IV-5. Typical profiles of the dependence of ΔG_{HX} on urea concentration. The points are derived from amide hydrogen/deuterium exchange experiments with PHS SNase and the lines are fitted using equations IV.6 and IV.7.

The values of ΔG_{HX} increase significantly in the presence of mannosylglycerate. Figure IV-6 shows that nearly all amide protons are stabilized, with a maximum increase of 2.3 kcal/mol for Y93 in the presence of 0.33 M mannosylglycerate. There is a weak correlation between the ΔG_{HX} of a given amide proton at zero denaturant concentration and $\Delta\Delta G_{HX}$ due to the presence of mannosylglycerate (R^2 = 0.43). There is no significant correlation between the extent of the global motion (m_2-value) and the value $\Delta\Delta G_{HX}$(global) (Figure IV-7).

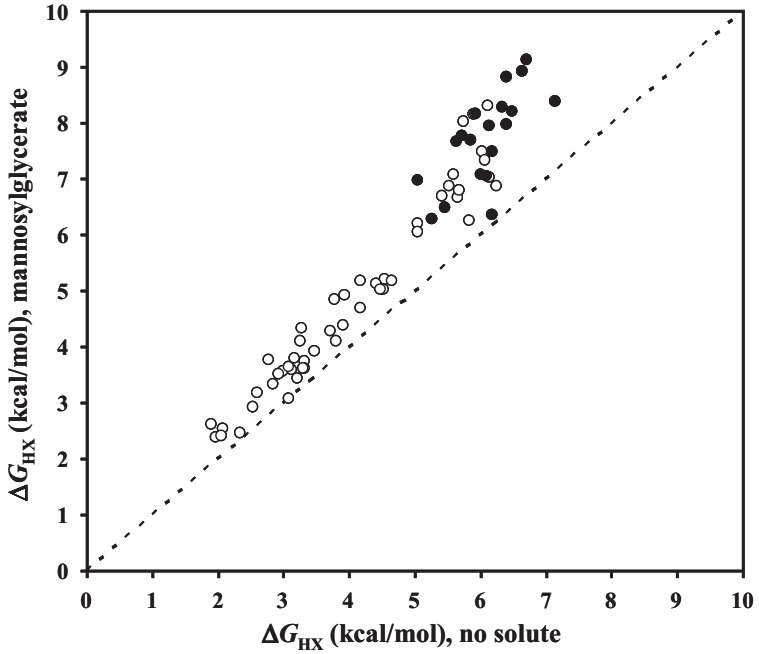


Figure IV-6. Free energy values (ΔG_{HX}) at zero denaturant concentration for the exchange of individual amide protons of a hyperstable SNase in the absence and presence of mannosylglycerate (0.33 M). The free energy values were obtained using equation IV.5. Symbols above the dashed line represent amide protons with increased ΔG_{HX} values in the presence of mannosylglycerate relative to no solute. Filled circles represent amide protons that exchange only through global opening motions (profile 1) while empty circles represent amide protons where exchange by local motions has a measurable contribution.

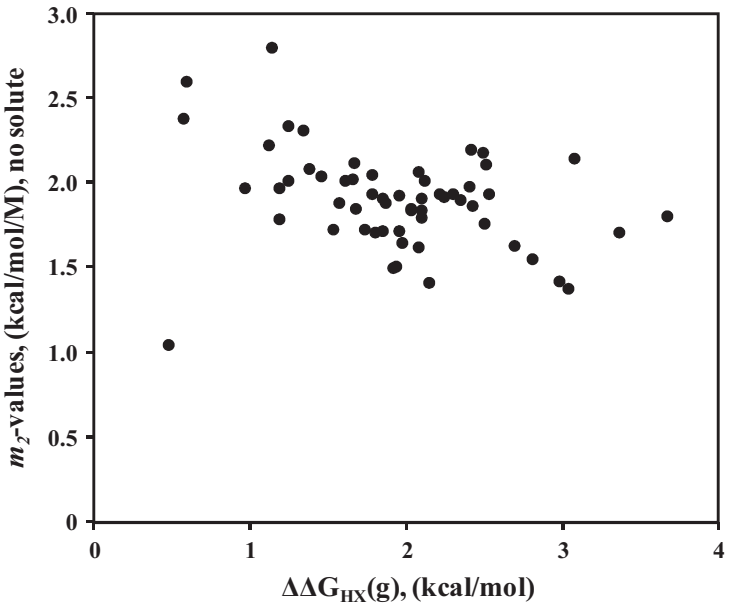


Figure IV-7. Free energy variation of the global opening motions as function of the m_2-value. The change in ΔG_{HX}(g) of the global opening motions induced by mannosylglycerate 0.33 M is plotted as a function of the respective m_2-value. $\Delta\Delta G_{HX}$(g) is obtained by subtracting the free energy values obtained by fitting the dependence of amide proton exchange on urea concentration in the presence of mannosylglycerate from those obtained in its absence.

Despite the very different ΔG_{HX} values exhibited by the amide protons of SNase, many eventually merge into common isotherms as the concentration of urea increases. This effect is illustrated by residues 129, 135 and 137 in Figure IV-5. Common ΔG_{HX} isotherms suggest cooperative unfolding and are characterised by similar m_2-values, ΔG_{HX}(g) values, and structural relatedness. The experimental data available to fit the parameters m_2 and ΔG_{HX}(g) is sometimes sparse, in which case small errors in the hydrogen exchange measurements produce significant variations in the values of the fitted parameters. Therefore, observing the whole trend of the ΔG_{HX} curve, instead of relying solely on the numerical values, seems more informative. In this way, we identify 45 amide protons as likely to be involved in ten

ΔG_{HX} isotherms (Figure IV-8). The presence of mannosylglycerate does not appear to cause any significant change in the compositions of the groups; groups four and six appear to merge (Figure IV.S-4), which seems to result from a greater increase of free energy values in group six relative to group four.

Large *m*-values tend to be associated with the more slowly exchanging amide protons, *i.e.*, those with higher ΔG_{HX}. However, 30% of the amides with ΔG_{HX} values greater than 5 kcal.mol^{-1} display profiles 2 or 3 with low initial *m*-values, precluding a general conclusion. The presence of mannosylglycerate has only marginal effects on the *m*-values and, in general, does not change the type of profile exhibited by a given amide proton. However, 6 amide protons (25, 67, 92, 99, 106, and 108) clearly shift from a two parameter model (equation IV.6) to a four parameter model (equation IV.7). Two of these amide protons are involved in an H-bond pair (Ile92 and Val99), two share the same H-bond acceptor (Gln106 and Leu108), and they are all located close to the hinge regions of the protein.

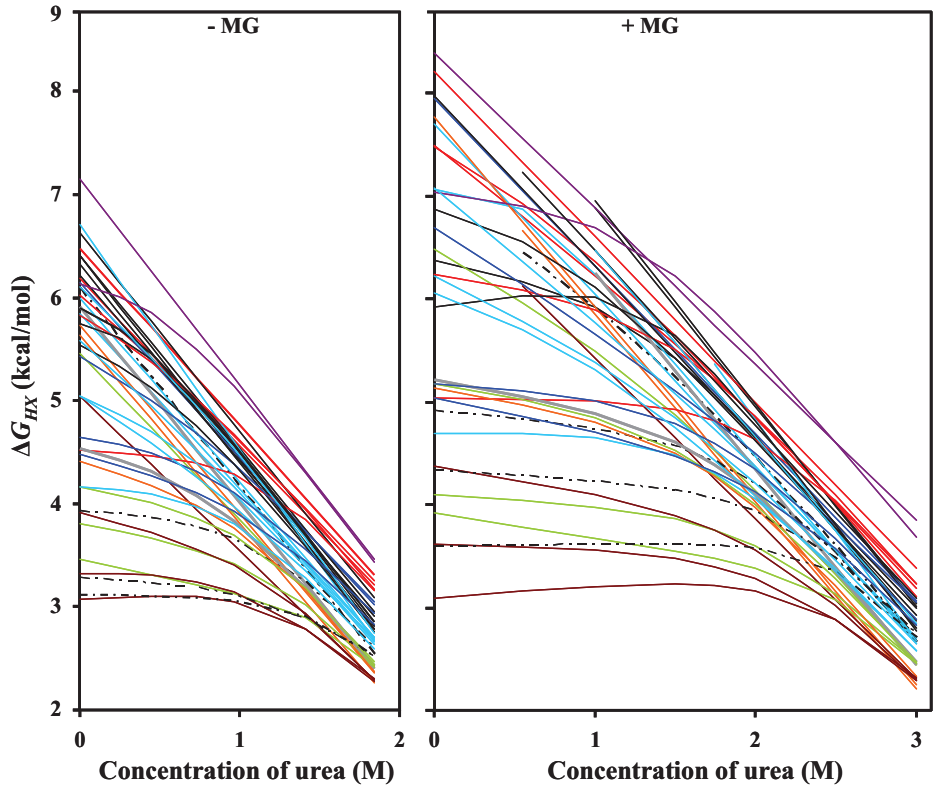


Figure IV-8. Partially unfolded units of SNase in the presence and absence of mannosylglycerate (0.33 M), characterized by merging ΔG_{HX}(g) isotherms. The ΔG_{HX}(g) isotherms reveal cooperative unfolding reactions of groups of amide protons promoted by urea in the absence and presence of mannosylglycerate (MG). The lines in the plot represent the best fit of the measured exchange rates using either equation IV.6 or IV.7. Each colour identifies amide protons involved in a particular partially unfolded unit: 1) purple for 36 and 37; 2) red for 97, 100, 102, 106, 107; 3) blue for 105, 128, 132, 134; 4) black for 24, 26, 30, 35, and 90 to 94; 5) light blue for 99, 101, 103, 104, 109, 110, 126, 130; 6) dashed line for 22, 23, 25, 27; 7) grey for 62 and 66, 8) green for 39, 108, 136, 139; 9) orange for 73, 74, 75; 10) 111, 129, 135, 137. The vertical axis has the same units in both plots.

DISCUSSION

This study complements our previous work on the effect of mannosylglycerate, a stabilizing solute accumulated by many (hyper)thermophilic organisms, on the backbone motions of a hyperstable staphylococcal nuclease variant. In that study we found a correlation between restriction of backbone dynamics and protein stabilization by mannosylglycerate. The analysis is now expanded to include fast side-chain motions, as determined by ^{13}C spin relaxation, and slower opening motions, as determined by H/D exchange measurements.

Fast dynamics of methyl bearing side-chains from ^{13}C relaxation measurements

Model-free analysis of the ^{13}C relaxation data obtained in the presence of the different solutes showed that model 5 (S^2_s, S^2_f, and τ_s) provided the best fit for the majority of the experimental data. This suggests a high degree of mobility of the protein side-chains and reorientation of the symmetry axis. The variation in the order parameter values between different fitted models has been noted previously and attributed to deviations from the extreme narrowing limit ($\tau_{e/s} << \tau_c$) (Lee et al. 1999; Wang et al. 1995). However, the majority of the methyl groups have $\tau_{e/s}$ and τ_c that differ by at least an order of magnitude and the variability may indicate that the models are oversimplified.

In contrast to the significant restriction of the backbone dynamics in SNase, there is no generalised restriction of side-chain fast dynamics as a function of mannosylglycerate stabilization (Figure IV-3). In other words, the side-chains remain essentially unperturbed in the presence of mannosylglycerate.

Since there is a theoretical correlation between changes in S^2 values and changes in conformational entropy (Yang and Kay 1996; Li et al. 1996), these

results suggest that mannosylglycerate reduces the entropy of the folded backbone and not of the side-chains. Furthermore, decreased unfolding entropy was proposed to be the origin of SNase A stabilization by mannosylglycerate below 65°C in a study based on differential scanning calorimetry and picosecond time-resolved fluorescence spectroscopy (Faria et al. 2004). These observations further support the involvement of motion restriction in protein stabilization by compatible solutes. They also underlie the importance of studying the effect of mannosylglycerate on the dynamics of the denatured state, using for example the Δ131Δ residue fragment that serves as a model for the denatured state of SNase.

Local and global opening motions in amide proton exchange

We followed the model of protein hydrogen exchange proposed by Linderström-Lang and developed by others (Hvidt and Nielsen 1966; Linderstrøm-Lang 1955; Englander and Kallenbach 1983) in which exchange occurs by opening motions of different extent that expose the hydrogen to the solvent. Because the amount of denaturant used here does not significantly change the population of folded protein (Figure IV-4), we are monitoring thinly populated unfolded states under normal conditions.

The majority of the monitored amide protons show evidence of exchange via local motions (profiles 2 and 3) which contrasts with the less stable wild type SNase in which nearly all amide protons (61 out of 64) exchange via global opening motions (Watson et al. 2007) (*i.e.*, profile 1). Amide protons with ΔG_{HX} values that do not change (m<0.1) in the range of urea concentrations tested here (profile 4) may indicate residual protection in the unfolded state of the protein (Bai et al. 1995). Gly88 may be an example of this as it was found in a native-like loop conformation in the Δ131Δ mutant, which is considered to be a good model of the unfolded structure of the wild

type SNase, (Zhang et al. 1997). Alternatively, the greatest urea concentration used here may not be sufficient for global motions to become dominant. As can be seen in Figure IV-6, mannosylglycerate restricts both local and global motions, though the range of effects on the global motions is somewhat larger.

Effect of mannosylglycerate on the ΔG_{HX} urea dependence profiles

Based on m_2-values, ΔG_{HX}(global), and structural relatedness we defined 10 groups of amide protons with merging unfolding isotherms and therefore possible cooperative unfolding. For some residues, the spread of the different global ΔG_{HX} isotherms is not sufficient to allow a definite assignment to one group or another. Nevertheless, it is clear that the relative position of the several ΔG_{HX} isotherms remains unchanged in the presence of mannosylglycerate (Figure IV-8). The *m*-values are also little affected by the presence of solute, which was also noted for trimethylamine *N*-oxide, another general stabilizer (Mello and Barrick 2003). These observations suggest that, despite the substantial increase in stability conferred by the solute, the intermediate states of the protein unfolding/folding pathway remain the same. Previous work from our group reached a similar conclusion but using different biophysical methods, namely picosecond time-resolved fluorescence and calorimetry (Faria et al. 2004).

Specific effects of mannosylglycerate on local motions

Many of the ΔG_{HX} curves obtained in the presence of mannosylglycerate (0.33 M) are virtually superimposable on those obtained in the absence of solute, with a shift of the ΔG_{HX} curve along the axis of urea concentration by about 1.2 M (see Figure IV-9 for examples). This overlap suggests that 0.33 M mannosylglycerate simply cancels the effect of 1.2 M urea. The superposition

of data points from lower concentrations of mannosylglycerate indicates that the cancellation occurs for all solute concentrations in the ratio of about 3.5 parts of urea to 1 of mannosylglycerate.

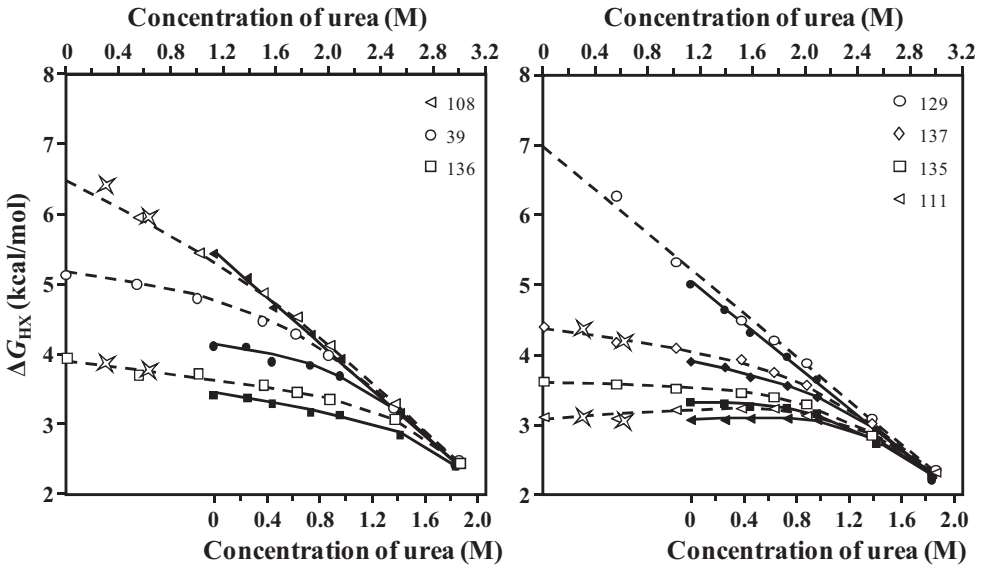


Figure IV-9. Comparison of the dependence of free energies of amide hydrogen exchange on urea concentration, determined at 35°C in the absence and presence of mannosylglycerate (0.33 M). The data obtained without mannosylglycerate is shifted along the concentration axis to show the similarity of the profiles. Empty symbols represent data obtained in the presence of 0.33 M mannosylglycerate and are plotted in relation to the axes above the panels; filled symbols represent data obtained without mannosylglycerate and are plotted in relation to the axes below the panels. The lines represent the best fit using equations IV.6 and IV.7. Star symbols represent selected ΔG_{HX} values obtained in the presence of 0.15 and 0.24 M mannosylglycerate and without urea, their positions on the concentration.

Mannosylglycerate (0.33 M) causes six amide protons to shift their ΔG_{HX} urea dependence profile from type 1 to types 2 or 3 (*e.g.,* residue 108 in Figure IV-9). This is a simple consequence of the general increase in ΔG_{HX}(g) values, such that local motions can be seen to contribute to the exchange.

However, the quality of cancellation of effects in 3.5:1 urea/mannosylglycerate mixtures is quite variable. Some amide protons show marked deviations, with increased ΔG_{HX}(local) values in the presence of mannosylglycerate, such as residue 39 in Figure IV-9. This does not fit the notion that stabilizers specifically restrict global motions, and the effect is not general.

To analyse the distribution of changes in ΔG_{HX}, we subtracted, for each amide proton, the ΔG_{HX} value obtained with 0.33 M mannosylglycerate and 1.2 M urea from the ΔG_{HX} value obtained without solutes and plotted them as a function of position in the structure (Figure IV-10). The width of the C_α trace of SNase in Figure IV-10 is proportional to this difference and clearly shows that the largest effects are associated with β-sheets. Thus, it appears that mannosylglycerate has a specific effect in restricting the local motions of amide groups in β-sheets.

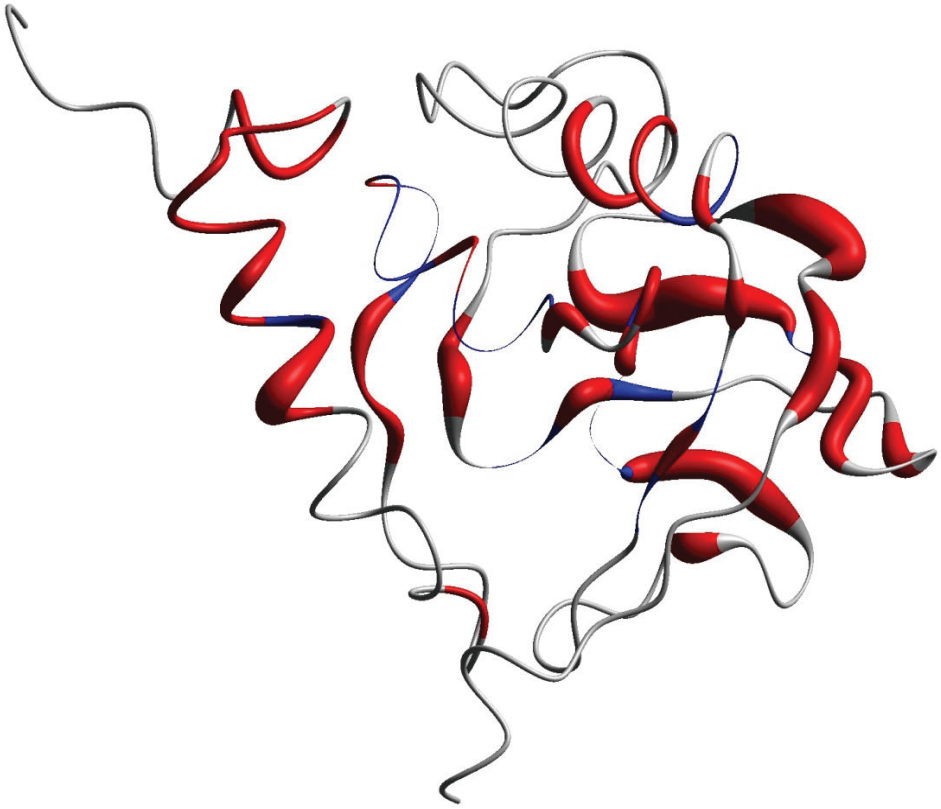


Figure IV-10. Distribution of deviations from cancellation of the effects of urea and mannosylglycerate on amide proton exchange. The Cα trace of SNase is shown with radii proportional to the differences between ΔG_{HX} values obtained in the presence of 1.2 M urea and 0.33 M mannosylglycerate and in the absence solutes. Residues with amide protons that exchange partly through local motions are coloured red, those exchanging only via global motions are in blue, and residues without data are in grey. The figure was created with MolMol (Koradi et al. 1996).

Further examples would be necessary to confirm that this is a stabilizing effect associated with a specific type of secondary structure rather than the response of one part of the SNase fold to an increase in effective surface tension across the whole protein. However, this effect goes some way to explain the observation that stabilization depends on the specific

solute/protein pair, to the extent that a normally stabilizing compound may have no effect on some proteins.

Concluding remarks

We have monitored the effect of mannosylglycerate, an effective protein stabilizer, on the motions of a hyperstable SNase variant using ^{13}C relaxation and hydrogen exchange measurements. This has extended earlier work both in timescales and accessible groups. Relaxation of ^{13}C methyl groups showed that side-chain motions in the sub-nanosecond time scale were little affected by the presence of mannosylglycerate (up to 0.35 M), in marked contrast to the behaviour observed for sub-nanosecond backbone motions which correlated strongly with changes in the melting temperature induced by different solutes (Pais et al. 2009).

On the other hand, studies on the slower time scale of amide hydrogen exchange measurements in the presence of urea revealed a new and unexpected phenomenon. Mannosylglycerate does not alter the folding/unfolding pathway of SNase, and restricts the global unfolding motions of the protein to a similar extent throughout the protein (Figure 6). The local motions, however, are restricted to varying extents, with some amides experiencing no change while others, particularly those in β-sheets, have changes in ΔG_{HX} that are comparable with those for global motions.

This work underlines the importance of studying the widest possible range of motions in exploring the link between protein dynamics and protein stabilization by compatible solutes. Furthermore, it suggests that mannosylglycerate may specifically restrict local motions in β-sheets, which may help to explain the variation in the degree of stabilization conferred on different proteins.

ACKNOWLEDGEMENTS AND WORK CONTRIBUTIONS

Dr. Manolis Matzapetakis was responsible for the work associated with the NMR determination of the SNase three-dimensional structure. This work was supported by Fundação para a Ciência e a Tecnologia (FCT), Portugal, Project PTDC/BIO/70806/2006 and grant number PEst-OE/EQB/LA0004/2011. The authors acknowledge Prof. Bertrand Garcia-Moreno, John Hopkins University, for providing the plasmid of the hyperstable SNase, and Dr. Luis Fonseca, ITQB, for valuable help with the statistics analysis and MatLab operation. Tiago M. Pais acknowledges FCT for grant support (SFRH/BD/42210/2007). The NMR spectrometers are part of The National NMR Network (REDE/1517/RMN/2005), supported by "Programa Operacional Ciência e Inovação (POCI) 2010" and FCT.

SUPPLEMENTARY MATERIAL

Table IV.S-1. Parameters extracted from the fits of hydrogen exchange rate measurements of SNase at 35°C (pD 5.2).

	no solute					with mannosylglycerate 0.33 M				
Residue	ΔG_{HX}	m_1	ΔG_{HX}(lc)	m_2	ΔG_{HX}(g)	ΔG_{HX}	m_1	ΔG_{HX}(lc)	m_2	ΔG_{HX}(g)
12	3.25	0.18	3.29	1.42	5.05	4.12	0.24	4.10	1.86	8.04
13	2.84	0.09	2.85	2.11	6.81	3.33	-0.14	3.36	0.84	5.32
15	3.79	0.21	3.80	1.93	5.75	4.85	0.18	4.85	1.78	7.53
16	2.80	0.10	2.77	3.96	10.17	3.79	0.07	3.76	1.78	7.97
21	1.95	-0.01	1.90	-	-	2.53	0.09	2.61	-	-
22	3.31	0.06	3.30	1.55	5.61	4.35	0.09	4.33	1.87	8.42
23	3.93	0.07	3.95	1.94	6.26	4.94	0.14	4.91	1.92	8.48
24	5.89	-0.18	6.16	2.01	6.55	8.15	-	-	1.69	8.15
25	6.09	-	-	1.94	6.09	*	1.00	7.11	1.95	8.62
26	5.53	0.35	5.68	2.02	6.47	6.83	0.22	6.95	1.73	8.12
27	3.12	0.01	3.12	2.11	6.71	3.58	-0.03	3.59	2.12	9.22
30	3.71	-0.02	3.73	1.90	6.72	4.27	-0.02	4.28	1.90	9.07
35	6.14	-	-	1.97	6.40	7.35	-	-	1.93	8.81
36	6.13	0.00	6.24	2.06	7.23	6.99	0.14	7.05	1.87	9.31
37	7.03	-	-	2.01	7.15	7.80	-	-	1.51	8.40
39	4.15	0.28	4.18	2.22	6.60	5.17	0.24	5.19	1.74	7.72
40	1.89	0.04	1.97	-	-	2.19	0.05	2.40	-	-
61	3.02	0.19	3.00	2.34	6.72	3.57	0.15	3.56	1.83	7.97
62	4.53	0.30	4.59	2.01	6.16	5.25	0.28	5.22	1.93	8.27
64	5.64	0.54	5.94	1.84	6.24	6.66	0.46	6.70	1.82	8.34
65	5.67	0.69	5.84	2.05	6.60	6.77	0.58	6.84	1.83	8.38
66	5.87	-	-	1.91	5.91	*	-	-	1.90	8.16
67	5.16	-	-	1.72	5.26	6.28	-0.37	6.44	1.71	7.21
69	2.66	0.01	2.54	-	-	2.74	0.02	2.92	-	-
71	2.14	-0.02	2.07	-	-	2.52	0.01	2.54	-	-
73	5.55	-	-	1.84	5.64	*	-	-	1.82	7.67
74	4.41	0.43	4.42	2.31	6.69	5.11	0.26	5.13	1.89	8.03

	no solute					with mannosylglycerate 0.33 M				
Residue	ΔG_{HX}	m_1	ΔG_{HX}(lc)	m_2	ΔG_{HX}(g)	ΔG_{HX}	m_1	ΔG_{HX}(lc)	m_2	ΔG_{HX}(g)
75	5.71	-	-	1.83	5.73	7.61	-	-	1.84	7.76
83	2.05	-	-	-	-	2.06	-0.06	2.04	-	-
87	2.57	0.09	2.59	-	-	3.18	-0.53	3.39	0.32	3.95
88	3.12	0.09	3.08	-	-	3.65	-0.04	3.65	1.51	7.65
89	3.17	0.05	3.16	2.18	6.45	3.80	0.08	3.79	2.11	8.94
90	6.24	-	-	1.93	6.32	*	-	-	1.80	8.28
91	6.09	-	-	1.88	6.41	*	-	-	1.66	7.98
92	6.13	-	-	1.80	6.22	6.39	0.28	6.38	1.96	8.69
93	5.72	-0.28	5.94	2.04	6.56	5.91	-0.26	5.92	2.08	9.01
94	6.59	-	-	1.93	6.63	*	-	-	1.98	8.94
95	2.05	0.00	2.05	-	-	2.40	0.08	2.42	-	-
97	4.52	0.06	4.53	2.12	7.12	5.05	0.00	5.04	1.84	8.79
99	5.91	-	-	1.79	6.00	7.07	-0.74	7.19	1.82	8.10
100	6.41	-	-	1.73	6.48	8.19	-	-	1.61	8.21
101	5.83	-	-	1.71	5.85	7.63	-	-	1.66	7.70
102	5.80	0.53	6.00	1.90	6.69	6.26	0.23	6.25	1.90	8.80
103	6.12	1.42	6.15	2.60	7.70	*	-	-	1.82	8.30
104	6.71	-	-	2.20	6.71	*	-	-	2.15	9.13
105	6.01	-	-	1.70	6.14	8.04	-	-	1.63	7.94
106	6.11	-	-	1.62	6.18	7.47	0.53	7.67	1.69	8.26
107	6.03	1.03	6.28	1.80	6.67	7.51	1.25	7.50	2.26	10.34
108	5.44	-	-	1.63	5.46	6.93	0.84	6.53	1.89	8.16
109	5.58	1.28	5.67	2.14	6.84	7.09	1.29	7.09	2.30	9.92
110	4.16	-0.02	4.19	1.84	6.02	4.70	-0.02	4.69	1.70	7.70
111	3.08	-0.14	3.08	1.72	5.57	3.12	-0.13	3.09	1.56	7.11
112	2.34	-0.08	2.34	4.79	12.35	2.47	-0.28	2.48	0.65	4.60
118	6.24	0.36	6.30	1.97	7.80	6.87	0.22	6.89	1.51	8.78
126	5.03	0.36	5.15	1.90	6.16	6.20	0.70	6.25	1.77	8.01
127	2.93	-0.05	2.93	2.38	7.38	3.50	-0.12	3.52	1.67	7.95
128	4.64	0.04	4.72	1.65	5.99	5.17	0.08	5.17	1.69	7.97
129	5.03	-	-	1.51	5.05	6.99	-	-	1.56	6.99

	no solute					with mannosylglycerate 0.33 M				
Residue	ΔG_{HX}	m_1	ΔG_{HX}(lc)	m_2	ΔG_{HX}(g)	ΔG_{HX}	m_1	ΔG_{HX}(lc)	m_2	ΔG_{HX}(g)
130	5.03	0.79	5.16	1.79	6.10	6.04	0.41	6.15	1.57	7.30
131	3.33	-0.03	3.34	1.41	5.51	3.77	-0.01	3.75	1.58	7.66
132	5.43	0.68	5.53	1.86	6.54	6.72	0.98	6.71	1.90	8.97
133	6.06	0.22	6.28	1.70	6.86	7.35	0.93	7.34	2.10	10.22
134	4.48	0.34	4.51	1.88	6.40	5.07	0.30	5.04	1.79	8.26
135	3.34	-0.13	3.36	1.50	5.10	3.62	0.03	3.62	1.55	7.02
136	3.82	0.24	3.81	1.97	6.19	4.13	0.09	4.11	1.61	7.38
137	3.91	0.37	3.93	2.08	6.26	4.41	0.25	4.38	1.76	7.64
139	3.45	0.33	3.46	2.80	7.97	3.95	0.24	3.92	2.15	9.11
140	3.22	-0.02	3.23	1.04	5.67	3.46	-0.04	3.45	0.89	6.15
141	3.29	0.05	3.30	1.37	5.82	3.59	0.08	3.61	1.89	8.86
Average	4.34	0.22	4.11	1.98	6.49	5.16	0.19	4.87	1.75	8.09

(a). ΔG_{HX}(lc) and ΔG_{HX}(g) are the free energies of the opening reactions associated with local (lc) and global (g) motions of the protein and are expressed in kcal/mol. (b). m_1 and m_2 are the slopes of the free energy dependence on urea concentration of the local and global motions, respectively, and are expressed in kcal/mol/M. (*) Free energy values not extracted because the respective exchange rates were to slow in the absence of urea to be properly determined within the experimental time.

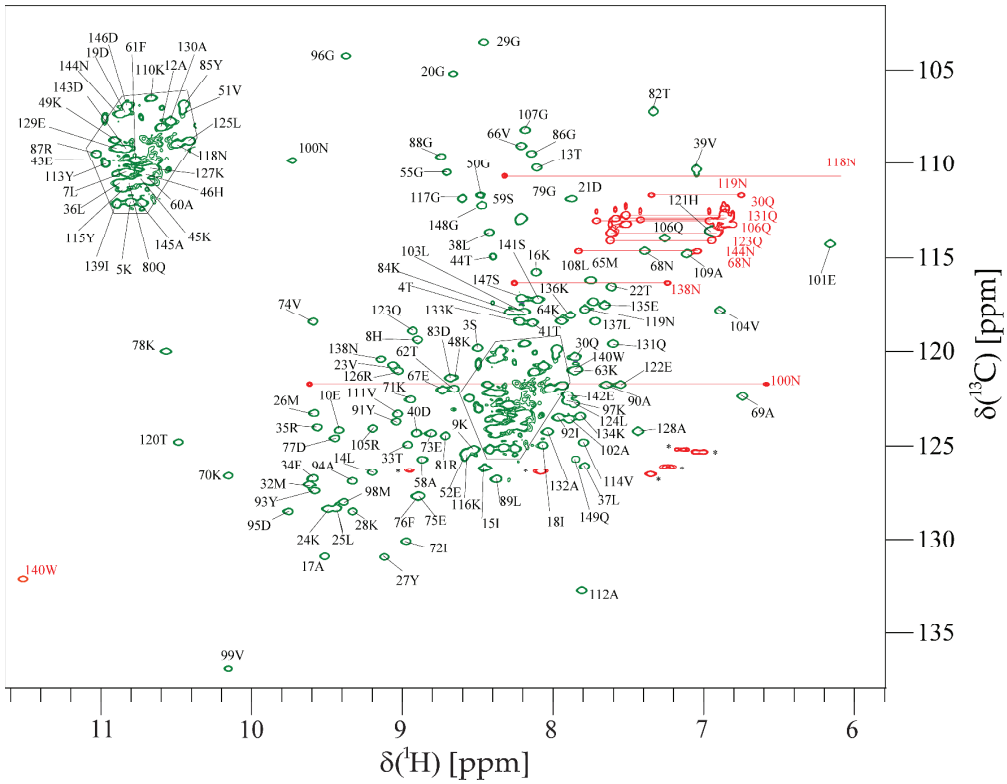


Figure IV.S-1. {^{1}H-^{15}N}-HSQC spectrum of the hyperstable SNase variant. Resonances in green correspond to backbone amides while red coloured belong to the respective sidechains. The sidechains with an asterisk correspond to folded resonances of Arg Nε– Hε (δ_{15N}~85ppm). The upper left insert shows the assignment of the peaks at the center of the spectrum.

Figure IV.S-2. Overlay of a representative model of the NMR bundle (model 6) shown in grey with the crystal structure of the PHS mutant of SNase (PDB ID: 1EY8) shown in purple. Only minor differences are observed, mainly in the unstructured loop (residues 42 to 53) and in the loops connecting sheets 2 and 3 (residues 83 to 86).

```
# Set the data pipe names (also the names of preset model-free models).
pipes = ['m0','m1', 'm2', 'm3', 'm4', 'm5', 'm9']

# Loop over the pipes.
for name in pipes:
    # Create the data pipe.
    pipe.create(name, 'mf')

    # Load the sequence.
    sequence.read('xfile.seq')

    # Load a PDB file.
    structure.read_pdb('structurePDBinDiffusionFrame.pdb')

    # Load spins
    structure.load_spins(spin_id='@CD1', combine_models=False, ave_pos=False)
    structure.load_spins(spin_id='@CG2', combine_models=False, ave_pos=False)
    structure.load_spins(spin_id='@CB', combine_models=False, ave_pos=False)
    structure.load_spins(spin_id='@CG1', combine_models=False, ave_pos=False)
    structure.load_spins(spin_id='@CD2', combine_models=False, ave_pos=False)

    # Set the spin name and then load the NH vectors.
    spin.name(spin_id='@CD1', name='CD1')
    structure.vectors(spin_id='@CD1', attached='H*', ave=False)
    spin.name(spin_id='@CG2', name='CG2')
    structure.vectors(spin_id='@CG2', attached='H*', ave=False)
    spin.name(spin_id='@CB', name='CB')
    structure.vectors(spin_id='@CB', attached='H*', ave=False)
    spin.name(spin_id='@CG1', name='CG1')
    structure.vectors(spin_id='@CG1', attached='H*', ave=False)
    spin.name(spin_id='@CD2', name='CD2')
    structure.vectors(spin_id='@CD2', attached='H*', ave=False)

    # Load the relaxation data.
   relax_data.read('R1',  '800', 800.333 * 1e6, 'R1_file', res_num_col=0, spin_name_col=2,
    data_col=3, error_col=4)
   relax_data.read('R2',  '800', 800.333 * 1e6, 'R2_file', res_num_col=0, spin_name_col=2,
    data_col=3, error_col=4)
   relax_data.read('NOE', '800', 800.333 * 1e6, 'Noe_file', res_num_col=0, spin_name_col=2,
    data_col=3, error_col=4)

    # Setup other values.
    diffusion_tensor.init((x, y, z, 0, 0, 0), param_types=2, fixed=True)
    value.set(1.07 * 1e-10, 'bond_length')
    value.set(0, 'csa')
    value.set('13C', 'heteronucleus')
    value.set('1H', 'proton')
```

```
    # Select the model-free model.
    model_free.select_model(model=name)

    # Minimise.
    grid_search(inc=11)
    minimise('newton')

    # Monte Carlo simulations.
    monte_carlo.setup(number=200)
    monte_carlo.create_data()
    monte_carlo.initial_values()
    minimise('newton')
    eliminate()
    monte_carlo.error_analysis()

    # Write the results.
    results.write(file='results', force=True)
    value.write('s2', file='xfile_out', dir=name, force=True)

# Save the program state.
state.save('save', force=True)
```

Figure IV.S-3. Script for the Model-free analysis using the RELAX software.

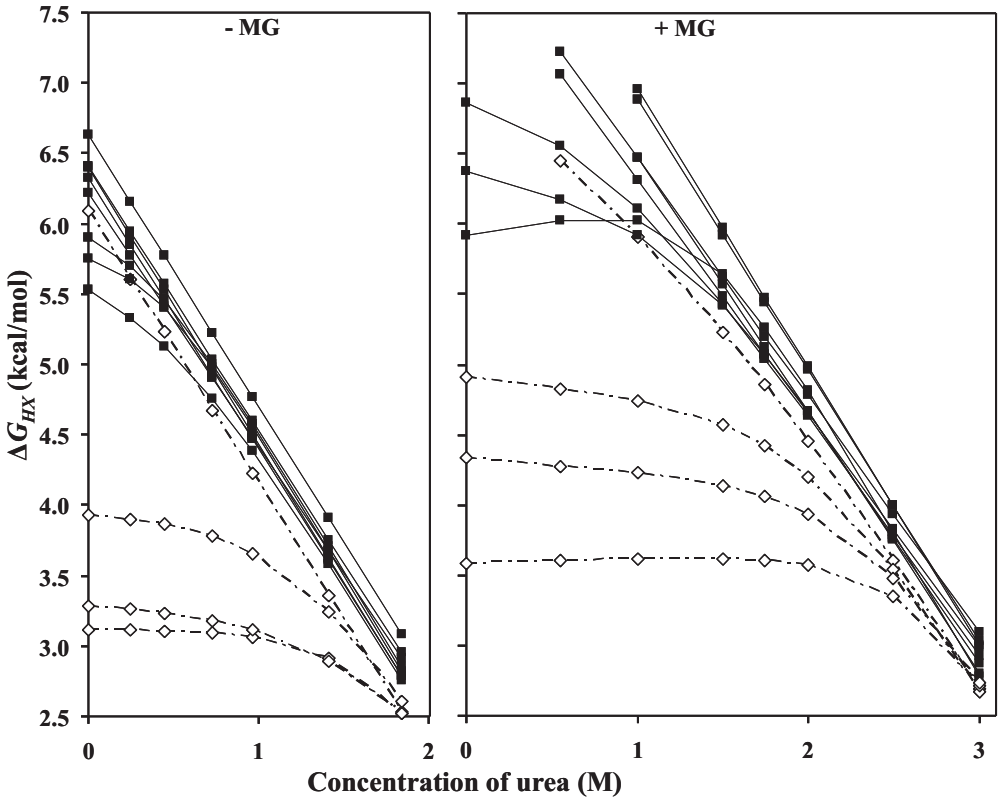


Figure IV.S-4. Detail of two partially unfolded units in the absence and presence of mannosylglycerate (0.33 M). The left panel shows free energy curves of amide protons for the protein in the absence of mannosylglycerate grouping in two partially unfolded units. The right panel shows the free energies of the same amide protons obtained in the presence of mannosylglycerate (0.33 M) which in this case appear to form a single partially unfolded unit. The vertical axis has the same units in both plots.

Chapter V

Concluding Discussion

CONTENTS

STRUCTURAL DETERMINANTS OF PROTEIN STABILIZATION BY SOLUTES

Studies performed *in vitro* have shown that there is no single compatible solute that can be considered the best protein stabilizer (Borges et al. 2002; Faria et al. 2008; Lamosa et al. 2000). Indeed, similar solutes can render very different extents of protection to the same protein, while the same solute can stabilize similar protein structures to very different extents – as shown by the studies with the different rubredoxins described in Chapter II. Additionally, the biological role of compatible solutes as thermal protectors of macromolecules *in vivo* is strongly suggested by the specificity of the solutes that some microorganisms produce in response to heat stress, which are clearly different from those generated under osmotic stress (Goncalves et al. 2003; Lamosa et

al. 1998; Silva et al. 1999). The specific nature of the magnitude of protein stabilization rendered by solutes *in vitro* most likely prevails *in vivo*, but other mechanisms of folding assistance, like chaperonin action (Liberek et al. 2008), may mask the contribution of solutes. In any case, the ability to accumulate specific compatible solutes is likely to be an advantage to the survival of microorganisms under heat stress.

As explained in Chapter I, the stability of the native protein results from the complex balance of folding and unfolding forces which, themselves, depend to a great extent on the interactions of the protein with the solvent. To understand how specific structural features of the protein may modulate the thermal protection conferred by compatible solutes we monitored the effect of two compatible solutes, diglycerol phosphate and mannosylglycerate, on a series of mutated rubredoxins characterized by single amino acid changes and by progressive amino acid deletions (Chapter II). Rubredoxin denaturation, as measured by the rate of iron release, is an irreversible process which implies that solute effects on the unfolded state should not contribute to the net protein stabilization. However, according to the Lumry-Eyring model of irreversible denaturation,

$$\mathbf{N \leftrightarrow U \rightarrow F} \qquad \textbf{[V.1]}$$

the native rubredoxin should be in equilibrium with a partially unfolded state that facilitates iron release. N represents the native state, U is the partially unfolded state, and F is the final unfolded state. Therefore, stabilizing solutes can still have an effect on both states and change the N ↔ U equilibrium. Mannosylglycerate and diglycerol phosphate are both negatively charged solutes, but they stabilize the several rubredoxin mutants to very different extents. While diglycerol phosphate reduces the rate of iron release of the wild-type rubredoxin by a factor of three, mannosylglycerate reduces it by a factor of only 1.3-fold. However, mannosylglycerate proved to be much more

efficient in increasing the stability of the deletion mutants than diglycerol phosphate. Surprisingly, one of the mutated rubredoxins (Δ23|29) was destabilized by the presence of diglycerol phosphate. These observations suggest different modes of action for these two negatively charged solutes. Moreover, in the particular case of diglycerol phosphate, the loop of the wild-type rubredoxin between residues 17 to 34 emerges as a potential structural determinant of the stabilization by this solute. Importantly, extreme shortening of this structural feature does not alter the configuration of the metal centre as revealed by the three dimensional structure of the mutant Δ17|29 determined in solution by NMR (Chapter II). In fact, the RMSD between the backbones of the structures of this mutant and the native rubredoxin from *D. gigas* is only 0.82 Å (excluding the residues that make up the shortened loop region). Further evidence of the role of this loop as a structural determinant in rubredoxin stabilization was provided by the changes in chemical shift observed in residues that were restricted to the loop region, or close neighbours, upon the addition of increasing amounts of diglycerol phosphate. These chemical shift variations suggest a preferential interaction of the solute with the loop region of the protein. A recent computational study using the structure of the *D. gigas* rubredoxin and diglycerol phosphate as model system (Micaelo et al. 2008) concludes that rubredoxin stabilization results from the solute opposing large scale deviations of the loop region – opening of this loop with exposure of the hydrophobic core is thought to be one of the first steps in the unfolding of this protein (Bonomi et al. 2000). The results of the computational study suggest that diglycerol phosphate molecules localize in the protein loop region via hydrogen bonding, which would explain the constraint of the motions of this structural feature. In our studies, we found indications of a tighter protein structure in the presence of the stabilizing solutes due to a strengthening of the hydrogen bond network.

However, no evidence for particular tightening at the loop region was found precluding a validation of the hypothesis put forward by Micaelo et al. (2008). On the other hand, the specific solute/protein interaction suggested by this computational study could help to explain the destabilization of mutant Δ23|29 in the presence of diglycerol phosphate. This is the only mutant with a net loss of two positively charged amino acids as a result of the deletion. Thus, the modified loop region concentrates four negatively charged residues:

Δ23|29 loop- **K^+.G.D^-.P.D^-.S.F.E^-.D^-.L.P**

Native loop- **K^+.G.D^-.P.D^-.S.**G.I.K^+.P.G.T.K^+.**F.E^-.D^-.L.P**

This cluster of negative charges may result in repulsive electrostatic interactions with the solute, which could lead to an increase in the opening motions of the loop, hence destabilization of the native rubredoxin.

Although the role of the loop region as a molecular determinant of rubredoxin stabilization by diglycerol phosphate seems well established, we lack the same evidence in the case of mannosylglycerate. It seems that the charged nature of the solutes *per se* does not determine the kind of interactions established with the protein surface. According to the current models, protein stabilization by solutes is largely explained on the basis of solute exclusion from the protein surface, either due to an increase in the water surface tension or to solvophobic effects, which are greater upon unfolding (Timasheff 2002; Auton et al. 2006; Auton and Bolen 2007; Arakawa et al. 2001). Exclusion is energetically unfavourable, resulting in a shift in the folding/unfolding equilibrium towards the native state. Although rubredoxins denature irreversibly, differential exclusion can still occur between the native state and the partly unfolded reversible state (*i.e.*, the N and U

states in the reaction scheme V.1), which would also shift the equilibrium to the native state. Importantly, solutes can still have affinity to specific protein groups resulting in interactions that will compete with the exclusion effects. Thus, the extent of exclusion depends on the balance between solute affinity to the protein and solute exclusion. In turn, this balance can vary throughout the surface of the protein according to the affinity of the solute to the exposed protein group. In this context, it is possible that mannosylglycerate exclusion from the water-protein interface is such that interactions of the solute with the loop region are weaker than in the case of diglycerol phosphate, which could be related with the different stabilization rendered by the two solutes. Despite the lack of knowledge about the degree of diglycerol phosphate exclusion from the protein surface, some considerations can be put forward. In chemical terms, diglycerol phosphate is a phosphate group bound to two glycerol moieties (Figure I-2). Glycerol alone does not increase the water surface tension and shows affinity for polar regions of the protein, however it is greatly excluded from the protein surface in the unfolded state due to unfavourable interactions with the backbone (Gekko and Timasheff 1981). Perhaps diglycerol phosphate partly retains these properties of the glycerol molecule, which can allow it to get closer to the protein surface than mannosylglycerate and, in this way, establish stronger electrostatic interactions with charged residues.

Naturally, there is no intention of making extrapolations or predictions for other systems based on these results, but these observations are important insofar as they go further in showing that specific interactions with selected structural features of the protein may have some relevance to the interpretation of protein stabilization data. On the other hand, a recent study suggests that the compaction of protein loops may explain the stabilizing effect of 0.05 M arginine/glutamic acid mixtures (Blobel et al. 2011). Using

small angle X-ray scattering and NMR, it was found that the flexible loops of three proteins (low molecular weight protein tyrosine phosphatase, FK-506 binding protein and chymotrypsinogen A) appeared to be selectively compacted upon the addition of a mixture of arginine and glutamic acid (Blobel et al. 2011). Thus, this study reports on a mode of action that resembles the one we proposed for the stabilization of rubredoxin mutants by diglycerol phosphate. It is possible that this represents a more general mechanism of protein stabilization by charged solutes. Furthermore, it has been reported that arginine forms clusters in solution, which may influence the arginine/protein interactions (Das et al. 2007) and eventually also the arginine/water interactions. This is another indication to reinforce the view that charged solutes display different modes of action that are different from those of neutral solutes. Finally, it would be interesting to construct mutants of the three proteins examined in the above-mentioned study (Blobel et al. 2011), with shortened loops, and analyze the impact of such mutations on the stabilization rendered by the arginine/glutamic acid mixture as well as by diglycerol phosphate and mannosylglycerate.

PROTEIN MOTIONS AND THEIR ROLE IN SOLUTE INDUCED STABILIZATION

The usual representation of the three dimensional structure of proteins is rather deceptive as it gives the impression of structural stiffness which, in most cases, is false. In fact, proteins display a number of internal motions covering different time scales and amplitudes. Therefore, a folded protein is more accurately described by an ensemble of conformations to reproduce the dynamic nature of the protein. Some of the conformers in the ensemble are energetically more favourable than others. The probability of a conformer is

related to its energy according to the Boltzmann distribution. Naturally, the least probable conformers are transient and will have little contribution to the determination of the protein structure. This does not necessarily mean that they are less important to the function of the protein. For example, molecular recognition and binding of ligands to proteins is frequently related to these excited native states (Sugase et al. 2007; Mulder et al. 2001; Iwahara and Clore 2006) by a mechanism called "selected fit" or conformational selection model (Tsai et al. 1999; Tsai et al. 2001b). In the selected-fit model, ligands bind to transient native state conformations and stabilize it, instead of binding to the ground-state population and inducing a conformational change (Figure V-1).

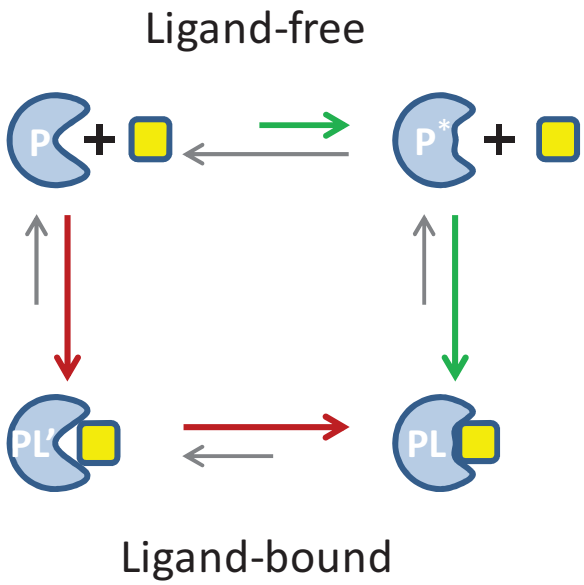


Figure V-1. Importance of protein motions in ligand binding via the "selected fit" model. The scheme represents the two models that are presently accepted for binding of the ligand to the protein. In the selected fit model (green arrows), the ligand (yellow square) binds to a transiently populated native conformation (P*) and stabilizes it (PL). Alternatively, in the induced fit model (red arrows), the ligand binds to the ground state conformation (P) and immediately induces a conformational change to a more stable state (PL).

Other important biological processes, such as enzyme catalysis and protein folding, have also been associated with these excited protein native states that are only formed due to the existence of protein motions (Cole and Loria 2003; Eisenmesser et al. 2002; Hill et al. 2000; Korzhnev et al. 2004). A classic example of the importance of protein motions to function is illustrated by haemoglobin in which rapid conformational rearrangements of the side-chains were found to be necessary to create the small crevices that allow oxygen to reach the haem iron (Perutz and Mathews 1966). More recently, a mutated T4 lysozyme was found to rapidly bind its substrate, despite steric blockage of the binding pocket (Mulder et al. 2001). Using NMR relaxation

techniques, the authors showed that the motions responsible for the excited native states (2.0 kcal/M) facilitate the entry of the ligand to the binding cavity.

It becomes clear that function and structure are intimately related with protein dynamics and that this relation is fairly well understood. On the other hand, the relationship between protein motions and protein stability is still a matter of great discussion. Indeed, the initial view linking protein rigidity with protein stability (Zavodszky et al. 1998; Tsai et al. 2001a) has been challenged (Kamerzell and Middaugh 2008). While the slowest and wider scale motions of the protein are broadly accepted as representing unfolding motion reactions, intermediate and fast time scale motions lack a clear relationship with protein stability (see Chapter I for a more comprehensive discussion of this matter).

The search for the relationship between protein dynamics and protein stability is not trivial mainly because a large number of factors which are difficult to isolate can contribute to protein stability. The most common approach to this subject has been to characterize protein dynamics in series of psychrophilic, mesophilic and thermophilic protein homologues. However, interpretation of results is complicated by the difficulty in filtering contributions from slight structural differences between protein homologues. Other issues include the precise quantification of dynamics at the different time scales and their respective contributions to protein stability. Thus, a fundamental question arises: which time scale motions are relevant to protein stability, if any? Is it the slow time scale in which motions appear to blend with the unfolding reaction? Is it the intermediate time scale in the micro to millisecond range which involves formation of excited native states? Or is it the fast sub-nanosecond time scale with an apparently limited impact over the

energy state? Finally, the possibility of a shared contribution of several time scales should also be considered.

As a result of these issues, it remains unclear whether more stable proteins are necessarily more rigid. However, if protein stability can be related with protein motions it is conceivable that protein stabilization by compatible solutes may also involve changes in protein flexibility. In such a context, the questions regarding the relevance of the different time scale motions also hold for protein stabilization. However, the use of compatible solutes makes it possible to manipulate the stability of the protein without altering the amino acid sequence or its structure. Therefore, it is possible to probe changes in dynamics as a function of changes in stability while using a single protein as object of study. Since solutes are not expected to significantly alter the protein native structures, this model system greatly facilitates data interpretation by avoiding the complications inherent to the structural variability of protein homologues.

The research data presented in the previous chapters of this dissertation (Chapters III and IV), underline the complex nature of the relationship between protein dynamics and protein stabilization. Indeed, even for a single time scale, contradictory conclusions can be drawn about the relevance of protein motions to stabilization, depending on the structural elements that are considered, *e.g.*, backbone or side-chains. These observation are, nevertheless, limited to the protein model used, a hyperstable staphylococcal nuclease variant (SNase), and a negatively charged solute (mannosylglycerate).

In the sub-nanosecond time scale, backbone motions of SNase (reflected by the generalized order parameters, S^2) were found to follow closely the changes in the melting temperature induced by several compounds, namely urea, mannosylglycerate, KCl and glycerol (Figure III-4.). Interestingly, the

restriction observed for the backbone motions in the presence of 0.35 M mannosylglycerate is equivalent to lowering the sample temperature by 5°C, *i.e.*, the change in the T_m induced by the solute. It may be a fortunate coincidence but, if meaningful, it would strengthen the inverse correlation between stability and backbone mobility observed for the other compounds tested. It should be noted that rigidification implies an entropic penalty for the protein. This penalty may be compensated by protein compaction, which reduces contact with water molecules and favours protein cohesive forces. In fact, binding of ligands to staphylococcal nuclease was shown to restrict fast backbone motions in several regions of the protein while increasing protein stability (Loh et al. 1993; Alexandrescu et al. 1996). Therefore, at least in the case of staphylococcal nuclease, the inverse correlation between backbone rigidification and protein stabilization induced by mannosylglycerate seems well founded (Figure V-2).

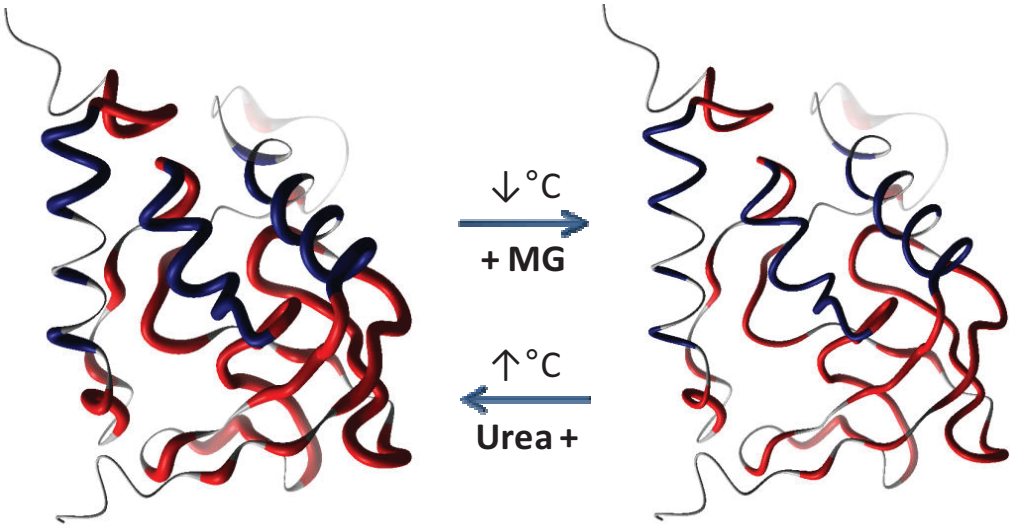


Figure V-2. Scheme representing changes in backbone dynamics (ΔS^2) of SNase induced by the presence of solutes or temperature shifts. The presence of mannosylglycerate (MG), or a reduction in temperature, restricts backbone motions (thinner backbone line); conversely, addition of a denaturing agent, urea, or a temperature increase, results in increased disorder.

With respect to side-chain sub-nanosecond motions, our results show that they are little affected by the presence of mannosylglycerate, or even low concentration of urea, in apparent contrast to what was observed for the backbone motions. In fact, most of the side-chains that were probed show a high degree of mobility including reorientation of the symmetry axis, *i.e.*, mobility about the bond that connects the methyl group to the residue side-chain. To the best of our knowledge there are no previous studies focused on monitoring fast side-chain dynamics in the presence of stabilizing compatible solutes. Studies using other protein models and other stabilizing solutes are necessary to assess the generality of the link between the restriction of backbone motions and protein stabilization. It is worth noting that side-chains, in particular non-polar ones, were shown to establish only weak interactions with stabilizing and destabilizing solutes while backbone groups establish

significant unfavourable interactions (Auton et al. 2011). Therefore, it is possible that the weak interaction between solutes and side-chains is not sufficient to produce any measurable impact on the fast motions of the protein side-chains.

In an intermediate time scale of protein motions, we probed chemical exchange of amide protons taking place in the range 20–500 milliseconds. Most of the eleven amide protons found to exchange within this time window were significantly slowed by the presence of mannosylglycerate. However, a similar reduction of the exchange rates was observed for most of the residues when glycerol was present in the sample: only Thr33 and Asn119 had exchange rates significantly slower in the presence of mannosylglycerate. Glycerol, at 0.60 M concentration, does not stabilize SNase and shows an equivalent viscosity to 0.25 M mannosylglycerate, making it a suitable control for viscosity effects on protein dynamics. Therefore, despite the significant reduction in the exchange rates induced by mannosylglycerate, the phenomenon does not appear to contribute significantly to the increase in protein stability observed in the presence of this solute. It should be noted, however, that the eleven residues probed represent only a small fraction of the protein and conclusions about the role of this time scale to protein stabilization by compatible solutes must await further experimental data. A greater number of probes of protein motions could be obtained by changing experimental parameters such as temperature and pH, which can bring chemical exchange of other amide protons to a time scale accessible to magnetization transfer NMR spectroscopy (*i.e.*, 20-500 milliseconds). Nevertheless, to better assess the importance of intermediate time scale motions it is necessary to have a broader picture of the effect of stabilizing solutes on the micro to millisecond time scale motions. Relaxation dispersion studies can be an interesting complementary approach since they probe

protein dynamics at faster time scales (microsecond range) than magnetization transfer experiments. This technique has the additional advantage of allowing thermodynamic information to be extracted about the exchange reaction (Palmer and Massi 2006; Baldwin and Kay 2009). Importantly, the effect of solutes on the microsecond time scale motions is still relatively unexplored.

Protein motions on the slow time scale (> millisecond) have often been studied by using hydrogen/deuterium exchange experiments (Wooll et al. 2000; Wang et al. 1995; Laurents et al. 2005; Krishna et al. 2004). Under certain conditions (see Chapter I for further detail), hydrogen exchange rates can be used to determine the free energy of the opening reaction (ΔG_{HX}) that enables the amide proton to exchange:

$$\Delta G_{\mathrm{HX}} = -RT\ ln\ K_{\mathrm{op}} = -RT \ln (k_{ex}/k_{ch}) \qquad [\mathbf{V.2}]$$

where k_{ex} is the measured exchange rate of an amide proton and k_{ch} is the exchange rate of the freely accessible amide as determined by other authors (Bai et al. 1993). Using this approach, several studies have shown that the free energy of the most slowly exchanging residues correlates reasonably well with the overall thermodynamic stability of proteins determined by other methodologies such as fluorescence and circular dichroism spectroscopies (Laurents et al. 2005; Fang et al. 2008; Bai et al. 1994). This correlation is thought to be due to the fact that the exchange reaction of the slowest exchanging hydrogens is controlled by transient global unfolding motions. However, hydrogens can also exchange via local opening motions. Global and local opening motions can be distinguished based on the urea dependence of the exchange rates. Local motions expose little or no new surface of the protein, resulting in exchange rates that are largely independent of denaturant

concentrations. Conversely, global motions expose significant new surface area and become increasingly favoured as the concentration of denaturant increases (Mayo and Baldwin 1993; Bai et al. 1995). In our studies, most of the amide protons of a hyperstable staphylococcal nuclease variant, exchange via local motions unlike the less stable wild type SNase where nearly all amide protons (61 out of 64) exchange via global motions (Watson et al. 2007). The presence of mannosylglycerate results in a generalised decrease of the exchange rates which is not matched by 0.60 M glycerol or 0.25 M KCl, indicating that the effect is not solely due to increased viscosity or ionic strength. Monitoring the urea dependence of hydrogen exchange rates showed that both global and local motions were significantly restricted by the solute. In apparent contrast, the solutes sucrose, glycerol and glycine were found to affect, mainly, amide groups exchanging via global motions (Wang et al. 1995; Foord and Leatherbarrow 1998; Calhoun and Englander 1985). The explanation for this difference may be related to the charged nature of mannosylglycerate which may result in different effects on the hydrogen exchange rates. In fact, the effects of salts on hydrogen exchange of bovine pancreatic trypsin inhibitor were suggested to be related to the variation of the electrostatic field at the exchanging site (Christoffersen et al. 1996).

A closer analysis of the ΔG_{HX}/urea dependence curves (Chapter IV) shows that, in the presence of mannosylglycerate, many amide protons have profiles that become virtually superimposable on those obtained in the absence of solute if the ΔG_{HX} curve is shifted along the axis of urea concentration by about 1.2 M. This overlap suggests that 0.33 M mannosylglycerate suppresses the effect of 1.2 M urea, that is, a 3.5:1 cancellation ratio. Trimethylamine *N*-oxide, a solute well known for counteracting the deleterious effects of urea (Qu and Bolen 2003; Yang and Kay 1996), shows a cancellation ratio of two. This means that for the same concentration of urea it would be necessary to

add twice as much trimethylamine *N*-oxide than mannosylglycerate to cancel out the disrupting effects of urea.

Not all amide protons display ΔG_{HX} curves that are superimposable. Some show significant deviations with increased ΔG_{HX}(local) values in the presence of mannosylglycerate, in particular at low urea concentrations (Figure IV-9). It turns out that the largest deviations of ΔG_{HX}(local) values were observed for the amide groups located in the β-sheets of the staphylococcal nuclease. We are not aware of another example in which the stabilizing effect of a solute is associated with protein motions of a specific secondary structural element. Despite the scarcity of studies, it is tempting to speculate that this localized effect may be one of the factors responsible for the dependence of the extent of stabilization on the particular solute/protein under consideration. Further examples would be necessary to test this hypothesis.

In summary, this thesis produced a considerable amount of data linking protein motions at different time scales to protein stabilization by mannosylglycerate. In particular, mannosylglycerate appears to stabilize staphylococcal nuclease by restricting fast backbone motions, but not side-chains, and by specifically constraining slow β-sheet local motions.

THE IMPORTANCE OF STUDYING PROTEIN STABILIZATION

The subject of protein stabilization by compatible solutes is, obviously intimately related to the study of protein intrinsic stability. We need to be aware of the forces that modulate protein stability in order to understand the effects underlying protein stabilization by solutes. Studies on the molecular mechanisms that govern protein stabilization by compatible solutes have been relatively successful in determining their impact on individual properties of the water-protein-solute system both in the folded and unfolded states of the

protein. For example, the effects of solutes on the protein surface (exclusion or binding), protein compactness, hydrogen bond network and the water structure (*e.g.,* changes in surface tension) have been well characterized for several protein-solute pairs (Arakawa and Timasheff 1985; Auton and Bolen 2005; Auton et al. 2006; Bolen 2001; Calhoun and Englander 1985; Faria et al. 2004; Tadeo et al. 2009; Timasheff 2002). However, the real contribution that each solute interaction makes to the thermodynamic stability of the protein has been much harder to unravel. Similar challenges exist in the field of protein intrinsic stability where the individual contribution of several forces have been accurately determined but the success in quantitatively predicting the intrinsic thermodynamic stability of the protein has been rather limited (Karaguler et al. 2004; Pace 1992; Jaenicke 1991; Casadio et al. 1995; Serrano et al. 1992). Therefore, understanding how proteins are stabilized by compatible solutes will surely contribute to the understanding of the forces that govern protein intrinsic stability.

Knowing how naturally occurring solutes stabilize proteins is not only a fundamental issue; it also has important biotechnological implications. Compounds that are able to stabilize the generally delicate nature of proteins may hold important applications in several fields of the biotechnology and pharmacology industries. Some classical solutes from mesophiles such as glycerol, trehalose and sucrose are frequently used to increase the shelf life of a diversity of foods and biomaterials. However, the discovery of compatible solutes that are exclusively or predominantly found in extremophilic organisms renewed the interest in the exploration of industrial applications for these small organic compounds. The field of application for general compatible solutes has, since then, expanded considerably with researchers finding promising uses for solutes like mannosylglycerate, mannosylgyceramide, ectoine, or hydroxyectoine in, i) DNA amplification, ii) protection of whole cells

and membranes during drying, storage and osmotic stress, iii) protection of skin against UV radiation, iv) inhibition of amyloid fibril formation, v) protection of plasmid DNA from hydroxyl radicals, and vi) protection of freeze-dried plasma components (Bakaltcheva et al. 2007; Lentzen and Schwarz 2006; Santos et al. 1998; Santos et al. 1999). Furthermore, stabilizing solutes have moved from mere additives to the main focus of the potential applications (*e.g.,* possible therapeutic drug for Alzheimer's disease). This shift in the application paradigm has stimulated the search for more applications as well as the research effort at the rationally design of organic molecules with improved properties (Faria et al. 2008).

Hence, the importance of further investigating the basis for the ability of compatible solutes to stabilize proteins is clearly apparent both in the fundamental and in the applied research fields.

References

Abragram, A. 1961. *The principles of nuclear magnetism*. Oxoford University Press, Inc.: New York.

Agashe, V.R., Shastry, M.C.R., and Udgaonkar, J.B. 1995. Initial hydrophobic collapse in the folding of barstar. *Nature* **377**:754-757.

Aguilar, C.F., Sanderson, I., Moracci,M., Ciaramella, M., Nucci, R., Rossi, M., and Pearl, L.H. 1997. Crystal structure of the beta-glycosidase from the hyperthermophilic archeon *Sulfolobus solfataricus*: resilience as a key factor in thermostability. *J. Mol. Biol.* **271**:789-802.

Alexandrescu, A.T., Jahnke, W., Wiltscheck, R., and Blommers, M.J. 1996. Accretion of structure in staphylococcal nuclease: an ^{15}N NMR relaxation study. *J. Mol. Biol.* **260**:570-587.

Allerhand, A., Doddrell, D., Glushko, V., Cochran, D.W., Wenkert, E., Lawson, P.J., and Gurd, F.R. 1971. Conformation and segmental motion of native and denatured ribonuclease A in solution. Application of natural-abundance carbon-13 partially relaxed Fourier transform nuclear magnetic resonance. *J. Am. Chem. Soc.* **93**:544-546.

Anfinsen, C.B. 1973. Principles that govern the folding of protein chains. *Science* **181**:223-230.

Arakawa, T., Ejima, D., Tsumoto, K., Obeyama, N., Tanaka, Y., Kita, Y., and Timasheff, S.N. 2007. Suppression of protein interactions by arginine: A proposed mechanism of the arginine effects. *Biophysical Chemistry* **127**:1-8.

Arakawa, T., Prestrelski, S.J., Kenney, W.C., and Carpenter, J.F. 2001. Factors affecting short-term and long-term stabilities of proteins. *Advanced Drug Delivery Reviews* **46**:307-326.

Arakawa, T. and Timasheff, S.N. 1984. Mechanism of protein salting in and salting out by divalent cation salts - Balance between hydration and salt binding. *Biochemistry* **23**:5912-5923.

Arakawa, T. and Timasheff, S.N. 1985. The stabilization of proteins by osmolytes. *Biophys. J.* **47**:411-414.

Auerbach, G., Ostendorp, R., Prade, L., Korndorfer, I., Dams, T., Huber, R., and Jaenicke, R. 1998. Lactate dehydrogenase from the hyperthermophilic bacterium *Thermotoga maritima*: the crystal structure at 2.1 angstrom resolution reveals strategies for intrinsic protein stabilization. *Structure with Folding & Design* **6**:769-781.

Auton, M. and Bolen, D.W. 2005. Predicting the energetics of osmolyte-induced protein folding/unfolding. *Proc. Natl. Acad. Sci. U. S. A.* **102**:15065-15068.

Auton, M. and Bolen, D.W. 2007. Application of the transfer model to understand how naturally occurring osmolytes affect protein stability. *Methods Enzymol.* **428**:397-418.

Auton, M. and Bolen, D.W. 2004. Additive transfer free energies of the peptide backbone unit that are independent of the model compound and the choice of concentration scale. *Biochemistry* **43**:1329-1342.

Auton, M., Ferreon, A.C.M., and Bolen, D.W. 2006. Metrics that differentiate the origins of osmolyte effects on protein stability: A test of the surface tension proposal. *Journal of Molecular Biology* **361**:983-992.

Auton, M., Holthauzen, L.M., and Bolen, D.W. 2007. Anatomy of energetic changes accompanying urea-induced protein denaturation. *Proc. Natl. Acad. Sci. U. S. A.* **104**:15317-15322.

Auton, M., Rosgen, J., Sinev, M., Holthauzen, L.M.F., and Bolen, D.W. 2011. Osmolyte effects on protein stability and solubility: A balancing act between backbone and side-chains. *Biophysical Chemistry* **159**:90-99.

Autumn, K., Sitti, M., Liang, Y.A., Peattie, A.M., Hansen, W.R., Sponberg, S., Kenny, T.W., Fearing, R., Israelachvili, J.N., and Full, R.J. 2002. Evidence for van der Waals adhesion in gecko setae. *Proc. Natl. Acad. Sci. U. S. A* **99**:12252-12256.

Bai, Y., Milne, J.S., Mayne, L., and Englander, S.W. 1993. Primary structure effects on peptide group hydrogen exchange. *Proteins* **17**:75-86.

Bai, Y., Milne, J.S., Mayne, L., and Englander, S.W. 1994. Protein stability parameters measured by hydrogen exchange. *Proteins* **20**:4-14.

Bai,Y., Sosnick, T.R., Mayne, L., and Englander, S.W. 1995. Protein folding intermediates: native-state hydrogen exchange. *Science* **269**:192-197.

Bakaltcheva, I., O'Sullivan, A.M., Hmel, P., and Ogbu, H. 2007. Freeze-dried whole plasma: Evaluating sucrose, trehalose, sorbitol, mannitol and glycine as stabilizers. *Thrombosis Research* **120**:105-116.

Baker, D. and Agard, D.A. 1994. Kinetics versus thermodynamics in protein-folding. *Biochemistry* **33**:7505-7509.

Baker, D., Sohl, J.L., and Agard, D.A. 1992. A protein-folding reaction under kinetic control. *Nature* **356**:263-265.

Baldwin, A.J. and Kay, L.E. 2009. NMR spectroscopy brings invisible protein states into focus. *Nat. Chem. Biol.* **5**:808-814.

Baldwin, R.L. 1986. Seeding protein folding. *Trends in Biochemical Sciences* **11**:6-9.

Barbar, E., Barany, G., and Woodward, C. 1996. Unfolded BPTI variants with a single disulfide bond have diminished non-native structure distance from the crosslink. *Fold. Des* **1**:65-76.

Batchelor, J.D., Olteanu, A., Tripathy, A., and Pielak, G.J. 2004. Impact of protein denaturants and stabilizers on water structure. *J. Am. Chem. Soc.* **126**:1958-1961.

Baxter, N.J. and Williamson, M.P. 1997. Temperature dependence of ^{1}H chemical shifts in proteins. *J. Biomol. NMR* **9**:359-369.

Bedard, S., Krishna, M.M.G., Mayne, L., and Englander, S.W. 2008a. Protein folding: Independent unrelated pathways or predetermined pathway with optional errors. *Proc. Natl. Acad. Sci.U. S. A.* **105**:7182-7187.

Bedard, S., Mayne, L.C., Peterson, R.W., Wand, A.J., and Englander, S.W. 2008b. The foldon substructure of staphylococcal nuclease. *J. Mol. Biol.* **376**:1142-1154.

Belev, T.N., Singh, M., and McCarthy, J.E. 1991. A fully modular vector system for the optimization of gene expression in *Escherichia coli*. *Plasmid* **26**:147-150.

Bell, G.S., Russell, R.J.M., Connaris, H., Hough, D.W., Danson, M.J., and Taylor, G.L. 2002a. Stepwise adaptations of citrate synthase to survival at life's extremes - From psychrophile to hyperthermophile. *European Journal of Biochemistry* **269**:6250-6260.

Bell, S., Hansen, S., and Buchner, J. 2002b. Refolding and structural characterization of the human p53 tumor suppressor protein. *Biophysical Chemistry* **96**:243-257.

Berger, A., Loewenstein, A., and Meiboom, S. 1959. Nuclear magnetic resonance study of the protolysis and ionization of *N*-methylacetamide. *J. Am. Chem. Soc.* **81**:62-67.

Bernado, P., Mylonas, E., Petoukhov, M.V., Blackledge, M., and Svergun, D.I. 2007. Structural characterization of flexible proteins using small-angle X-ray scattering. *J. Am. Chem. Soc.* **129**:5656-5664.

Blake, P.R., Park, J.B., Bryant, F.O., Aono, S., Magnuson, J.K., Eccleston, E., Howard, J.B., Summers, M.F., and Adams, M.W.W. 1991. Determinants of protein hyperthermostability - purification and amino-acid sequence of rubredoxin from the hyperthermophilic archaebacterium *Pyrococcus furiosus* and secondary structure of the zinc adduct by Nmr. *Biochemistry* **30**:10885-10895.

Blobel, J., Brath, U., Bernado, P., Diehl, C., Ballester, L., Sornosa, A., Akke, M., and Pons, M. 2011. Protein loop compaction and the origin of the effect of arginine and glutamic acid mixtures on solubility, stability and transient oligomerization of proteins. *Eur. Biophys. J.*

Bolen, D.W. 2001. Protein stabilization by naturally occurring osmolytes. *Methods Mol. Biol.* **168**:17-36.

Bonomi, F., Burden, A.E., Eidsness, M.K., Fessas, D., Iametti, S., Kurtz, D.M., Mazzini, S., Scott, R.A., and Zeng, Q.D. 2002. Thermal stability of the [Fe(SCys)(4)] site in *Clostridium pasteurianum* rubredoxin: contributions of the local environment and Cys ligand protonation. *Journal of Biological Inorganic Chemistry* **7**:427-436.

Bonomi, F., Fessas, D., Iametti, S., Kurtz, D.M., and Mazzini, S. 2000. Thermal stability of *Clostridium pasteurianum* rubredoxin: Deconvoluting the contributions of the metal site and the protein. *Protein Science* **9**:2413-2426.

Borges, N., Ramos, A., Raven, N.D., Sharp, R.J., and Santos, H. 2002. Comparative study of the thermostabilizing properties of mannosylglycerate and other compatible solutes on model enzymes. *Extremophiles.* **6**:209-216.

Bougault, C.M., Eidsness, M.K., and Prestegard, J.H. 2003. Hydrogen bonds in rubredoxins from mesophilic and hyperthermophilic organisms. *Biochemistry* **42**:4357-4372.

Brennan, L., Turner, D.L., Messias, A.C., Teodoro, M.L., Legall, J., Santos, H., and Xavier, A.V. 2000. Structural basis for the network of functional cooperativities in cytochrome *c* (3) from *Desulfovibrio gigas*: solution structures of the oxidised and reduced states. *J. Mol. Biol.* **298**:61-82.

Briand, J. and Ernst, R.R. 1991. Computer-optimized homonuclear TOCSY experiments with suppression of cross relaxation. *Chemical Physics Letters* **185**:276-285.

Brown, A.D. 1976. Microbial Water Stress. *Bacteriological Reviews* **40**:803-846.

Brown, A.D. and Simpson, J.R. 1972. Water relations of sugar-tolerant yeasts: the role of intracellular polyols. *J. Gen. Microbiol.* **72**:589-591.

Butterwick, J.A., Patrick, L.J., Astrof, N.S., Kroenke, C.D., Cole, R., Rance, M., and Palmer, A.G., III 2004. Multiple time scale backbone dynamics of homologous thermophilic and mesophilic ribonuclease HI enzymes. *J. Mol. Biol.* **339**:855-871.

Calhoun, D.B. and Englander, S.W. 1985. Internal protein motions, concentrated glycerol, and hydrogen exchange studied in myoglobin. *Biochemistry* **24**:2095-2100.

Cambillau, C. and Claverie, J.M. 2000. Structural and genomic correlates of hyperthermostability. *J. Biol. Chem.* **275**:32383-32386.

Casadio, R., Compiani, M., Fariselli, P., and Vivarelli, F. 1995. Predicting free energy contributions to the conformational stability of folded proteins from the residue sequence with radial basis function networks. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **3**:81-88.

Cavagnero, S., Debe, D.A., Zhou, Z.H., Adams, M.W.W., and Chan, S.I. 1998a. Kinetic role of electrostatic interactions in the unfolding of hyperthermophilic and mesophilic rubredoxins. *Biochemistry* **37**:3369-3376.

Cavagnero, S., Zhou, Z.H., Adams, M.W.W., and Chan, S.I. 1998b. Unfolding mechanism of rubredoxin from *Pyrococcus furiosus*. *Biochemistry* **37**:3377-3385.

Chen, J., Lu, Z., Sakon, J., and Stites, W.E. 2000. Increasing the thermostability of staphylococcal nuclease: implications for the origin of protein thermostability. *J. Mol. Biol.* **303**:125-130.

Cheng, Y., Oldfield, C.J., Meng, J., Romero, P., Uversky, V.N., and Dunker,A.K. 2007. Mining alpha-helix-forming molecular recognition features with cross species sequence alignments. *Biochemistry* **46**:13468-13477.

Cho, M.K., Kim, H.Y., Bernado, P., Fernandez, C.O., Blackledge, M., and Zweckstetter, M. 2007. Amino acid bulkiness defines the local conformations and dynamics of natively unfolded alpha-synuclein and tau. *J. Am. Chem. Soc.* **129**:3032-3033.

Christoffersen, M., Bolvig, S., and Tuchsen, E. 1996. Salt effects on the amide hydrogen exchange of bovine pancreatic trypsin inhibitor. *Biochemistry* **35**:2309-2315.

Clore, G.M., Szabo, A., Bax, A., Kay, L.E., Driscoll, P.C., and Gronenborn, A.M. 1990. Deviations from the simple 2-parameter model-free approach to the interpretation of ^{15}N Nuclear Magnetic Relaxation of proteins. *J. Am. Chem. Soc.* **112**:4989-4991.

Cohen, F.E. and Kelly, J.W. 2003. Therapeutic approaches to protein-misfolding diseases. *Nature* **426**:905-909.

Cole, R. and Loria, J.P. 2003. FAST-Modelfree: a program for rapid automated analysis of solution NMR spin-relaxation data. *J. Biomol. NMR* **26**:203-213.

Cooper, A. 1999. Thermodynamics of protein folding and stability. In *Proteins: a comprehensive treatise*. (ed. G Allen), pp 217-270. JAI Press Inc.

Cordier, F., Caffrey, M., Brutscher, B., Cusanovich, M.A., Marion, D., and Blackledge, M. 1998. Solution structure, rotational diffusion anisotropy and local backbone dynamics of *Rhodobacter capsulatus* cytochrome *c* 2. *J. Mol. Biol.* **281**:341-361.

Crowhurst, K.A. and Forman-Kay, J.D. 2003. Aromatic and methyl NOEs highlight hydrophobic clustering in the unfolded state of an SH3 domain. *Biochemistry* **42**:8687-8695.

D'Amico, S., Marx, J.C., Gerday, C., and Feller, G. 2003. Activity-stability relationships in extremophilic enzymes. *J. Biol. Chem.* **278**:7891-7896.

d'Auvergne, E.J. and Gooley, P.R. 2008b. Optimisation of NMR dynamic models II. A new methodology for the dual optimisation of the model-

free parameters and the Brownian rotational diffusion tensor. *J. Biomol. NMR* **40**:121-133.

d'Auvergne, E.J. and Gooley, P.R. 2008a. Optimisation of NMR dynamic models I. Minimisation algorithms and their performance within the model-free and Brownian rotational diffusion spaces. *J. Biomol. NMR* **40**:107-119.

da Costa, M.S., Santos, H., and Galinski, E.A. 1998. An overview of the role and diversity of compatible solutes in Bacteria and Archaea. *Adv. Biochem. Eng Biotechnol.* **61**:117-153.

Daniel, R.M. and Danson, M.J. 2001. Assaying activity and assessing thermostability of hyperthermophilic enzymes. *Methods Enzymol.* **334**:283-293.

Das, U., Hariprasad, G., Ethayathulla, A.S., Manral, P., Das, T.K., Pasha, S., Mann, A., Ganguli, M., Verma, A.K., Bhat, R., Chandrayan, S.K., Ahmed, S., Sharma, S., Kaur, P., Singh, T.P., and Srinivasan, A. 2007. Inhibition of protein aggregation: supramolecular assemblies of arginine hold the key. *Plos One* **2**.

Dauter, Z., Wilson, K.S., Sieker, L.C., Moulis, J.M., and Meyer, J. 1996. Zinc- and iron-rubredoxins from *Clostridium pasteurianum* at atomic resolution: A high-precision model of a ZnS4 coordination unit in a protein. *Proc. Natl. Acad. Sci. U. S. A.* **93**:8836-8840.

Davis-Searles, P.R., Saunders, A.J., Erie, D.A., Winzor, D.J., and Pielak, G.J. 2001. Interpreting the effects of small uncharged solutes on protein-folding equilibria. *Annu. Rev. Biophys. Biomol. Struct.* **30**:271-306.

Dedmon, M.M., Christodoulou, J., Wilson, M.R., and Dobson, C.M. 2005. Heat shock protein 70 inhibits alpha-synuclein fibril formation via preferential binding to prefibrillar species. *J. Biol. Chem.* **280**:14733-14740.

Deechongkit, S., Nguyen, H., Powers, E.T., Dawson, P.E., Gruebele, M., and Kelly, J.W. 2004. Context-dependent contributions of backbone hydrogen bonding to beta-sheet folding energetics. *Nature* **430**:101-105.

Doan-Nguyen, V. and Loria, J.P. 2007. The effects of cosolutes on protein dynamics: the reversal of denaturant-induced protein fluctuations by trimethylamine *N*-oxide. *Protein Sci.* **16**:20-29.

Dobson, C.M., Lian, L.-Y., Redfield, C., and Topping, K.D. 1986. Measurement of hydrogen exchange rates using 2D NMR spectroscopy. *J. Magn Reson.* **69**:201-209.

Dosset, P., Hus, J.C., Blackledge, M., and Marion, D. 2000. Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data. *J. Biomol. NMR* **16**:23-28.

Eidsness, M.K., Richie, K.A., Burden, A.E., Kurtz, D.M., and Scott, R.A. 1997. Dissecting contributions to the thermostability of *Pyrococcus furiosus* rubredoxin: beta-sheet chimeras. *Biochemistry* **36**:10406-10413.

Eisenmesser, E.Z., Bosco, D.A., Akke, M., and Kern, D. 2002. Enzyme dynamics during catalysis. *Science* **295**:1520-1523.

Englander, S.W. and Kallenbach, N.R. 1983. Hydrogen-exchange and structural dynamics of proteins and nucleic-acids. *Quarterly Reviews of Biophysics* **16**:521-655.

Englander, S.W. and Mayne, L. 1992. Protein folding studied using hydrogen-exchange labeling and two-dimensional NMR. *Annu. Rev. Biophys. Biomol. Struct.* **21**:243-265.

Englander, S.W., Mayne, L., Bai, Y., and Sosnick, T.R. 1997. Hydrogen exchange: the modern legacy of Linderstrøm-Lang. *Protein Sci.* **6**:1101-1109.

Englander, S.W., Mayne, L., and Krishna, M.M.G. 2007. Protein folding and misfolding: mechanism and principles. *Quarterly Reviews of Biophysics* **40**:287-326.

Englander, S.W., Sosnick, T.R., Englander, J.J., and Mayne, L. 1996. Mechanisms and uses of hydrogen exchange. *Curr. Opin. Struct. Biol.* **6**:18-23.

Enninga, J., Mounier, J., Sansonetti, P., and Tran Van, N.G. 2005. Secretion of type III effectors into host cells in real time. *Nat. Methods* **2**:959-965.

Fang, X., Cui, Q., Tong, Y., Feng, Y., Shan, L., Huang, L., and Wang, J. 2008. A stabilizing alpha/beta-hydrophobic core greatly contributes to hyperthermostability of archaeal [P62A]Ssh10b. *Biochemistry* **47**:11212-11221.

Faria, T.Q., Lima, J.C., Bastos, M., Macanita, A.L., and Santos, H. 2004. Protein stabilization by osmolytes from hyperthermophiles: effect of mannosylglycerate on the thermal unfolding of recombinant nuclease a from *Staphylococcus aureus* studied by picosecond time-resolved fluorescence and calorimetry. *J. Biol. Chem.* **279**:48680-48691.

Faria, T.Q., Mingote, A., Siopa, F., Ventura, R., Maycock, C., and Santos, H. 2008. Design of new enzyme stabilizers inspired by glycosides of hyperthermophilic microorganisms. *Carbohydr. Res.* **343**:3025-3033.

Farrow, N.A., Muhandiram, R., Singer, A.U., Pascal, S.M., Kay, C.M., Gish, G., Shoelson, S.E., Pawson, T., Forman-Kay, J.D., and Kay, L.E. 1994. Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by ^{15}N NMR relaxation. *Biochemistry* **33**:5984-6003.

Fersht, A.R. 1987. The hydrogen-bond in molecular recognition. *Trends in Biochemical Sciences* **12**:301-304.

Finkelstein, A.V. and Ptitsyn, O.B. 2002. *Protein physics: a course of lectures*. Academic Press.

Foord, R.L. and Leatherbarrow, R.J. 1998. Effect of osmolytes on the exchange rates of backbone amide protons in proteins. *Biochemistry* **37**:2969-2978.

Fuxreiter, M., Tompa, P., Simon, I., Uversky, V.N., Hansen, J.C., and Asturias, F.J. 2008. Malleable machines take shape in eukaryotic transcriptional regulation. *Nature Chemical Biology* **4**:728-737.

Ganguly, D. and Chen, J.H. 2011. Topology-based modeling of intrinsically disordered proteins: Balancing intrinsic folding and intermolecular interactions. *Proteins* **79**:1251-1266.

Garrett, J.B., Kretz, K.A., O'Donoghue, E., Kerovuo, J., Kim, W., Barton, N.R., Hazlewood, G.P., Short, J.M., Robertson, D.E., and Gray, K.A. 2004. Enhancing the thermal tolerance and gastric performance of a microbial phytase for use as a phosphate-mobilizing monogastric-feed supplement. *Appl. Environ. Microbiol.* **70**:3041-3046.

Gekko, K. and Timasheff, S.N. 1981. Thermodynamic and kinetic examination of protein stabilization by glycerol. *Biochemistry* **20**:4677-4686.

Gershenson, A., Schauerte, J.A., Giver, L., and Arnold, F.H. 2000. Tryptophan phosphorescence study of enzyme flexibility and unfolding in laboratory-evolved thermostable esterases. *Biochemistry* **39**:4658-4665.

Gillespie, J.R. and Shortle, D. 1997. Characterization of long-range structure in the denatured state of staphylococcal nuclease. II. Distance restraints from paramagnetic relaxation and calculation of an ensemble of structures. *J. Mol. Biol.* **268**:170-184.

Goncalves, L.G., Huber, R., da Costa, M.S., and Santos, H. 2003. A variant of the hyperthermophile *Archaeoglobus fulgidus* adapted to grow at high salinity. *Fems Microbiology Letters* **218**:239-244.

Guntert, P., Braun, W., and Wuthrich, K. 1991. Efficient computation of three-dimensional protein structures in solution from nuclear magnetic resonance data using the program Diana and the supporting programs Caliba, Habas and Glomsa. *J. Mol. Biol.* **217**:517-530.

Güntert, P., Mumenthaler, C., and Wuthrich, K. 1997. Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.* **273**:283-298.

Hernandez, G., Jenney, F.E. Jr., Adams, M.W., and LeMaster, D.M. 2000. Millisecond time scale conformational flexibility in a hyperthermophile protein at ambient temperature. *Proc. Natl. Acad. Sci. U. S. A.* **97**:3166-3170.

Herrmann, T.G., Güntert, P., and Wüthrich, K. 2002a. Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J. Mol. Biol.* **319**:209-227.

Herrmann, T.G., Güntert, P., and Wüthrich, K. 2002b. Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS. *J. Biomol. NMR.* **24**:171-189.

Hill, R.B., Bracken, C., DeGrado, W.F., and Palmer, A.G. 2000. Molecular motions and protein folding: Characterization of the backbone dynamics and folding equilibrium of alpha D-2 using ^{13}C NMR spin relaxation. *J. Am.Chem. Soc.* **122**:11610-11619.

Horne, J., d'Auvergne, E.J., Coles, M., Velkov, T., Chin, Y., Charman, W.N., Prankerd, R., Gooley, P.R., and Scanlon, M.J. 2007. Probing the flexibility of the DsbA oxidoreductase from *Vibrio cholerae* - a ^{15}N - ^{1}H heteronuclear NMR relaxation analysis of oxidized and reduced forms of DsbA. *J. Mol. Biol.* **371**:703-716.

Hu, C.Y., Kokubo, H., Lynch, G.C., Bolen, D.W., and Pettitt, B.M. 2010. Backbone additivity in the transfer model of protein solvation. *Protein Science* **19**:1011-1022.

Hvidt, A. 1964. A discussion of the pH dependence of the hydrogen-deuterium exchange of proteins. *C. R. Trav. Lab Carlsberg.* **34**:299-317.

Hvidt, A. and Nielsen, S.O. 1966. Hydrogen exchange in proteins. *Adv. Protein Chem.* **21**:287-386.

Hwang, T.L., Mori, S., Shaka, A.J., and van Zijl, P.C. 1997. Application of phase-modulated CLEAN chemical EXchange spectroscopy (CLEANEX-PM) to detect water-protein proton exchange and intermolecular NOEs. *J. Am. Chem. Soc.* **119**:6203-6204.

Hwang, T.L., van Zijl, P.C., and Mori, S. 1998. Accurate quantitation of water-amide proton exchange rates using the phase-modulated CLEAN chemical EXchange (CLEANEX-PM) approach with a Fast-HSQC (FHSQC) detection scheme. *J. Biomol. NMR* **11**:221-226.

Iakoucheva, L.M., Brown, C.J., Lawson, J.D., Obradovic, Z., and Dunker, A.K. 2002. Intrinsic disorder in cell-signaling and cancer-associated proteins. *J. Mol. Biol.* **323**:573-584.

Iwahara, J. and Clore, G.M. 2006. Detecting transient intermediates in macromolecular binding by paramagnetic NMR. *Nature* **440**:1227-1230.

Jacobs, M.D. and Fox, R.O. 1994. Staphylococcal nuclease folding intermediate characterized by hydrogen-exchange and NMR spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **91**:449-453.

Jaenicke, R. 1996. How do proteins acquire their three-dimensional structure and stability? *Naturwissenschaften* **83**:544-554.

Jaenicke, R. 1991. Protein stability and molecular adaptation to extreme conditions. *European Journal of Biochemistry* **202**:715-728.

Jaenicke, R. and Bohm, G. 1998. The stability of proteins in extreme environments. *Curr. Opin. Struct. Biol.* **8**:738-748.

Jaenicke, R., Schurig, H., Beaucamp, N., and Ostendorp, R. 1996. Structure and stability of hyperstable proteins: glycolytic enzymes from hyperthermophilic bacterium *Thermotoga maritima*. *Adv. Protein Chem.* **48**:181-269.

Jaravine, V.A., Rathgeb-Szabo, K., and Alexandrescu, A.T. 2000. Microscopic stability of cold shock protein A examined by NMR native state hydrogen exchange as a function of urea and trimethylamine *N*-oxide. *Protein Science* **9**:290-301.

Jaswal, S.S., Sohl, J.L., Davis, J.H., and Agard, D.A. 2002. Energetic landscape of alpha-lytic protease optimizes longevity through kinetic stability. *Nature* **415**:343-346.

Jeener, J., Meier, B.H., Bachmann, P., and Ernst, R.R. 1979. Investigation of exchange processes by two-dimensional NMR spectroscopy. *J. Chem. Phys.* **71**:4546-4553.

Kamerzell, T.J. and Middaugh, C.R. 2008. The complex inter-relationships between protein flexibility and stability. *J. Pharm. Sci.* **97**:3494-3517.

Karaguler, N.G., Sessions, R.B., Moreton, K.M., Clarke, A.R., and Holbrook, J.J. 2004. Estimating the energetic contribution of hydrogen bonding to the stability of *Candida methylica* formate dehydrogenase by using double mutant cycle. *Biotechnol. Lett.* **26**:1137-1140.

Kaushik, J.K. and Bhat, R. 1999. A mechanistic analysis of the increase in the thermal stability of proteins in aqueous carboxylic acid salt solutions. *Protein Science* **8**:222-233.

Kay, L.E. 2005. NMR studies of protein structure and dynamics. *Journal of Magnetic Resonance* **173**:193-207.

Kay, L.E., Muhandiram, D.R., Wolf, G., Shoelson, S.E., and Forman-Kay, J.D. 1998. Correlation between binding and dynamics at SH2 domain interfaces. *Nature Structural Biology* **5**:156-163.

Kay, L.E., Torchia, D.A., and Bax, A. 1989. Backbone dynamics of proteins as studied by ^{15}N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry* **28**:8972-8979.

Knapp, S., deVos, W.M., Rice, D., and Ladenstein, R. 1997. Crystal structure of glutamate dehydrogenase from the hyperthermophilic eubacterium *Thermotoga maritima* at 3.0 angstrom resolution. *J. Mol. Biol.***267**:916-932.

Koradi, R., Billeter, M., and Wuthrich, K. 1996. MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**:51-32.

Korzhnev, D.M. and Kay, L.E. 2008. Probing invisible, low-populated states of protein molecules by relaxation dispersion NMR spectroscopy: an application to protein folding. *Acc. Chem. Res.* **41**:442-451.

Korzhnev, D.M., Salvatella, X., Vendruscolo, M., Di Nardo, A.A., Davidson, A.R., Dobson, C.M., and Kay, L.E. 2004. Low-populated folding

intermediates of Fyn SH3 characterized by relaxation dispersion NMR. *Nature* **430**:586-590.

Krishna, M.M.G., Hoang, L., Lin,Y., and Englander, S.W. 2004. Hydrogen exchange methods to study protein folding. *Methods* **34**:51-64.

Krishnamurthy, H., Munro, K., Yan, H., and Vieille, C. 2009. Dynamics in *Thermotoga neapolitana* adenylate kinase: ^{15}N relaxation and hydrogen-deuterium exchange studies of a hyperthermophilic enzyme highly active at 30°C. *Biochemistry* **48**:2723-2739.

Kudou, M., Shiraki, K., Fujiwara, S., Imanaka, T., and Takagi, M. 2003. Prevention of thermal inactivation and aggregation of lysozyme by polyamines. *European Journal of Biochemistry* **270**:4547-4554.

Lamosa, P., Brennan, L., Vis, H., Turner, D.L., and Santos, H. 2001. NMR structure of *Desulfovibrio gigas* rubredoxin: a model for studying protein stabilization by compatible solutes. *Extremophiles.* **5**:303-311.

Lamosa, P., Burke, A., Peist, R., Huber, R., Liu, M.Y., Silva, G., Rodrigues-Pousada, C., LeGall, J., Maycock, C., and Santos, H. 2000. Thermostabilization of proteins by diglycerol phosphate, a new compatible solute from the hyperthermophile *Archaeoglobus fulgidus*. *Appl. Environ. Microbiol.* **66**:1974-1979.

Lamosa, P., Martins, L.O., da Costa, M.S., and Santos, H. 1998. Effects of temperature, salinity, and medium composition on compatible solute accumulation by *Thermococcus* spp. *Applied and Environmental Microbiology* **64**:3591-3598.

Lamosa, P., Turner, D.L., Ventura, R., Maycock, C., and Santos, H. 2003. Protein stabilization by compatible solutes. Effect of diglycerol phosphate on the dynamics of *Desulfovibrio gigas* rubredoxin studied by NMR. *European Journal of Biochemistry.* **270**:4606-4614.

Larsen, T.M., Benning, M.M., Wesenberg, G.E., Rayment, I., and Reed, G.H. 1997. Ligand-induced domain movement in pyruvate kinase: Structure of the enzyme from rabbit muscle with Mg^{2+}, K^{+}, and L-phospholactate

at 2.7 angstrom resolution. *Archives of Biochemistry and Biophysics* **345**:199-206.

Laskowski, R.A., Rullmann, J.A.C., MacArthur, M.W., Kaptein, R., and Thornton, J.M. 1996. AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *Journal of Biomolecular NMR* **8**:477-486.

Laurents, D.V., Scholtz, J.M., Rico, M., Pace, C.N., and Bruix, M. 2005. Ribonuclease Sa conformational stability studied by NMR-monitored hydrogen exchange. *Biochemistry* **44**:7644-7655.

Lebbink, J.H.G., Knapp, S., van der Oost, J., Rice, D., Ladenstein, R., and De Vos, W.M. 1999. Engineering activity and stability of *Thermotoga maritima* glutamate dehydrogenase. II: Construction of a 16-residue ion-pair network at the subunit interface. *J. Mol. Biol.* **289**:357-369.

Lee, A.L., Flynn, P.F., and Wand, A.J. 1999. Comparison of H-2 and C-13 NMR relaxation techniques for the study of protein methyl group dynamics in solution. *J. Am. Chem. Soc.* **121**:2891-2902.

LeMaster, D.M., Tang, J., Paredes, D.I., and Hernandez, G. 2005. Enhanced thermal stability achieved without increased conformational rigidity at physiological temperatures: spatial propagation of differential flexibility in rubredoxin hybrids. *Proteins* **61**:608-616.

Lentzen, G. and Schwarz, T. 2006. Extremolytes: natural compounds from extremophiles for versatile applications. *Applied Microbiology and Biotechnology* **72**:623-634.

Li, Z., Raychaudhuri, S., and Wand, A.J. 1996. Insights into the local residual entropy of proteins provided by NMR relaxation. *Protein Sci.* **5**:2647-2650.

Liberek, K., Lewandowska, A., and Zietkiewicz, S. 2008. Chaperones in control of protein disaggregation. *EMBO Journal* **27**:328-335.

Lietzow, M.A., Jamin, M., Jane Dyson, H.J., and Wright, P.E. 2002. Mapping long-range contacts in a highly unfolded protein. *J. Mol. Biol.* **322**:655-662.

Linderstrøm-Lang, K. 1958 Deuterium exchange and protein structure. In *Symposium on protein structure*, NeubergerA, ed. London.

Linderstrøm-Lang, K. 1955. The pH-dependence of the deuterium exchange of insulin. *Biochim. Biophys. Acta* **18**:308.

Lipari, G. and Szabo, A. 1982a. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules 1. Theory and range validity. *J. Am. Chem. Soc* **104**:4546-4559.

Lipari, G. and Szabo, A. 1982b. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules 2. Analysis of experimental results. *J. Am. Chem. Soc* **104**:4559-4570.

Liu, Y. and Bolen, D.W. 1995. The peptide backbone plays a dominant role in protein stabilization by naturally occurring osmolytes. *Biochemistry* **34**:12884-12891.

Liu, Y. and Sturtevant, J.M. 1996. The observed change in heat capacity accompanying the thermal unfolding of proteins depends on the composition of the solution and on the method employed to change the temperature of unfolding. *Biochemistry* **35**:3059-3062.

Loh, S.N., Prehoda, K.E., Wang, J.F., and Markley, J.L. 1993. Hydrogen-exchange in unligated and ligated staphylococcal nuclease. *Biochemistry* **32**:11022-11028.

Lopez, C.J., Fleissner, M.R., Guo, Z.F., Kusnetzow, A.K., and Hubbell, W.L. 2009. Osmolyte perturbation reveals conformational equilibria in spin-labeled proteins. *Protein Science* **18**:1637-1652.

Lumry, R. and Eyring, H. 1954. Conformation changes of proteins. *Journal of Physical Chemistry* **58**:110-120.

Mandel, A.M., Akke, M., and Palmer, A.G., III 1995. Backbone dynamics of *Escherichia coli* ribonuclease HI: correlations with structure and function in an active enzyme. *J. Mol. Biol.* **246**:144-163.

Markley, J.L., Horsley, W.J., and Klein, M.P. 1971. Spin-lattice relaxation measurements in slowly relaxing complex spectra. *The J. Chem. Phys.* **55**:3604-3605.

Mayo, S.L. and Baldwin, R.L. 1993. Guanidinium chloride induction of partial unfolding in amide proton exchange in RNase A. *Science* **262**:873-876.

McDonald, I.K. and Thornton, J.M. 1994. Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* **238**:777-793.

Mello, C.C. and Barrick, D. 2003. Measuring the stability of partly folded proteins using TMAO. *Protein Sci.* **12**:1522-1529.

Micaelo, N.M., Victor, B.L., and Soares, C.M. 2008. Protein thermal stabilization by charged compatible solutes: Computational studies in rubredoxin from *Desulfovibrio gigas*. *Proteins* **72**:580-588.

Miller, W.G. and Goebel, C.V. 1968. Dimensions of protein random coils. *Biochemistry* **7**:3925-3935.

Mittermaier, A.K. and Kay, L.E. 2009. Observing biological dynamics at atomic resolution using NMR. *Trends in Biochemical Sciences* **34**:601-611.

Molday, R.S., Englander, S.W., and Kallen, R.G. 1972. Primary structure effects on peptide group hydrogen exchange. *Biochemistry* **11**:150-158.

Mori, S., Abeygunawardana, C., van Zijl, P.C., and Berg, J.M. 1996. Water exchange filter with improved sensitivity (WEX II) to study solvent-exchangeable protons. Application to the consensus zinc finger peptide CP-1. *J. Magn Reson. B* **110**:96-101.

Mori, S., Wardana, C.A., Johnson, M.O., and van Zijl, P.C. 1995. Improved sensitivity of HSQC spectra of exchanging protons at short interscan

delays using a new Fast-HSQC (FHSQC) detection scheme that avoids water saturation. *J. Magn Reson. B* **108**:94-95.

Mozhaev, V.V. 1993. Mechanism-based strategies for protein thermostabilization. *Trends Biotechnol.* **11**:88-95.

Mulder, F.A., Mittermaier, A., Hon, B., Dahlquist, F.W., and Kay, L.E. 2001. Studying excited states of proteins by NMR spectroscopy. *Nat. Struct. Biol.* **8**:932-935.

Murphy, K.P. 1995. Noncovalent forces important to the conformational stability of protein structures. *Methods Mol. Biol.* **40**:1-34.

Nederveen, A.J., Doreleijers, J.F., Vranken, W., Miller, Z., Spronk, C.A., Nabuurs, S.B., Güntert, P., Livny, M., Markley, J.L., Nilges, M., Ulrich, E.L., Kaptein, R., Bonvin, A.M. 2005. RECOORD: a recalculated coordinate database of 500+ proteins from the PDB using restraints from the BioMagResBank. *Proteins* **59**:662-672.

Nicholson, L.K., Kay, L.E., Baldisseri, D.M., Arango, J., Young, P.E., Bax, A., and Torchia, D.A. 1992. Dynamics of methyl groups in proteins as studied by proton-detected ^{13}C NMR spectroscopy. Application to the leucine residues of staphylococcal nuclease. *Biochemistry* **31**:5253-5263.

Nodet, G., Salmon, L., Ozenne, V., Meier, S., Jensen, M.R., and Blackledge, M. 2009. Quantitative description of backbone conformational sampling of unfolded proteins at amino acid resolution from NMR residual dipolar couplings. *J. Am. Chem. Soc.* **131**:17908-17918.

Nölting, B. 1999. Physical interactions that determine the properties of proteins. In *Protein folding kinetics: biophysical methods*. (ed. Nölting), pp 17-25. Springer-Verlag: Berlin.

Noronha, M., Gerbelova, H., Faria,T.Q., Lund, D.N., Smith, D.A., Santos, H., and Macanita, A.L. 2010. Thermal unfolding kinetics of ubiquitin in the microsecond-to-second time range probed by Tyr-59 fluorescence. *J. Phys. Chem. B* **114**:9912-9919.

Noronha, M., Gerbelova, H., Faria, T.Q., Sampaio, M.M., Rudolph, R., Macanita, A.L., and Santos, H. 2009. Enhancing the fluorescence of tyr-59 in ubiquitin by blocking proton transfer. *Phys. Chem. Chem. Phys.* **11**:3580-3583.

Nozaki, Y. and Tanford, C. 1964. Hydrogen Ion titration curves of rabbit gamma-globulin + its papain-cleaved fragments. *Federation Proceedings* **23**:453-&.

Nozaki, Y. and Tanford, C. 1963. The solubility of amino acids and related compounds in aqueous urea solutions. *J. Biol. Chem.* **238**:4074-4081.

Nozaki, Y. and Tanford, C. 1965. The solubility of amino acids and related compounds in aqueous ethylene glycol solutions. *J. Biol. Chem.* **240**:3568-3575.

Nozaki, Y. and Tanford, C. 1970. The solubility of amino acids, diglycine, and triglycine in aqueous guanidine hydrochloride solutions. *J. Biol. Chem.* **245**:1648-1652.

Orekhov, V.Y., Nolde, D.E., Golovanov, A.P., Korzhnev, D.M., and Arseniev, A.S. 1995. Processing of heteronuclear NMR relaxation data with the new software DASHA. *Applied Magnetic Resonance* **9**:581-588.

Pace, C.N. 1992. Contribution of the hydrophobic effect to globular protein stability. *J. Mol. Biol.* **226**:29-35.

Pace, C.N., Shirley, B.A., McNutt, M., and Gajiwala, K. 1996. Forces contributing to the conformational stability of proteins. *Faseb Journal* **10**:75-83.

Pais, T.M., Lamosa, P., dos Santos, W., Legall, J., Turner, D.L., and Santos, H. 2005. Structural determinants of protein stabilization by solutes: the importance of the hairpin loop in rubredoxins. *FEBS J.* **272**:999-1011.

Pais, T.M., Lamosa, P., Garcia-Moreno, B., Turner, D.L., and Santos, H. 2009. Relationship between protein stabilization and protein rigidification induced by mannosylglycerate. *J. Mol. Biol.* **394**:237-250.

Palackal, N., Brennan, Y., Callen, W.N., Dupree, P., Frey, G., Goubet, F., Hazlewood, G.P., Healey, S., Kang, Y.E., Kretz, K.A., Lee, E., Tan, X.Q., Tomlinson, G.L., Verruto, J., Wong, V.W.K., Mathur, E.J., Short, J.M., Robertson, D.E., and Steer, B.A. 2004. An evolutionary route to xylanase process fitness. *Protein Science* **13**:494-503.

Palmer, A.G., III 2001. NMR probes of molecular dynamics: overview and comparison with other techniques. *Annu. Rev. Biophys. Biomol. Struct.* **30**:129-155.

Palmer, A.G., III, Hochstrasser, R.A., Millar, D.P., Rance, M., and Wright, P.E. 1993. Characterization of amino-acid side-chain dynamics in a zinc-finger peptide using ^{13}C NMR spectroscopy and time-resolved fluorescence spectroscopy. *J. Am. Chem. Soc.* **115**:6333-6345.

Palmer, A.G. and Massi, F. 2006. Characterization of the dynamics of biomacromolecules using rotating-frame spin relaxation NMR spectroscopy. *Chem. Rev.* **106**:1700-1719.

Palmer, A.G., Rance, M., and Wright, P.E. 1991. Intramolecular motions of a zinc finger DNA-binding domain from Xfin characterized by proton-detected natural abundance ^{12}C heteronuclear NMR spectroscopy. *J. Am. Chem. Soc.* **113**:4371-4380.

Pan, H., Barbar, E., Barany, G., and Woodward, C. 1995. Extensive nonrandom structure in reduced and unfolded bovine pancreatic trypsin inhibitor. *Biochemistry* **34**:13974-13981.

Perutz, M.F. and Mathews, F.S. 1966. An X-ray study of azide methaemoglobin. *J. Mol. Biol.* **21**:199-&.

Petsko, G.A. 2001. Structural basis of thermostability in hyperthermophilic proteins, or "there's more than one way to skin a cat". *Methods Enzymol.* **334**:469-478.

Pfeil, W., Gesierich, U., Kleemann, G.R., and Sterner, R. 1997. Ferredoxin from the hyperthermophile *Thermotoga maritima* is stable beyond the boiling point of water. *J. Mol. Biol.* **272**:591-596.

Piotto, M., Saudek, V., and Sklenar, V. 1992. Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR* **2**:661-665.

Pletneva, E.V., Gray, H.B., and Winkler, J.R. 2005. Many faces of the unfolded state: conformational heterogeneity in denatured yeast cytochrome *c*. *J. Mol. Biol.* **345**:855-867.

Qu, Y., Bolen, C.L., and Bolen, D.W. 1998. Osmolyte-driven contraction of a random coil protein. *Proc. Natl. Acad. Sci. U. S. A.* **95**:9268-9273.

Qu, Y. and Bolen, D.W. 2003. Hydrogen exchange kinetics of RNase A and the urea:TMAO paradigm. *Biochemistry* **42**:5837-5849.

Ramos, A., Raven, N., Sharp, R.J., Bartolucci, S., Rossi, M., Cannio, R., Lebbink, J., Van Der, O.J., De Vos, W.M., and Santos, H. 1997. Stabilization of enzymes against thermal stress and freeze-drying by mannosylglycerate. *Appl. Environ. Microbiol.* **63**:4020-4025.

Ramos, C.H. and Baldwin, R.L. 2002. Sulfate anion stabilization of native ribonuclease A both by anion binding and by the Hofmeister effect. *Protein Sci.* **11**:1771-1778.

Ratcliff, K. and Marqusee, S. 2010. Identification of residual structure in the unfolded state of ribonuclease H1 from the moderately thermophilic *Chlorobium tepidum*: comparison with thermophilic and mesophilic homologues. *Biochemistry* **49**:5167-5175.

Razvi, A. and Scholtz, J.M. 2006. Lessons in stability from thermophilic proteins. *Protein Sci.* **15**:1569-1578.

Richarz, R., Nagayama, K., and Wuthrich, K. 1980. Carbon-13 nuclear magnetic resonance relaxation studies of internal mobility of the polypeptide chain in basic pancreatic trypsin inhibitor and a selectively reduced analogue. *Biochemistry* **19**:5189-5196.

Roberts, M.F. 2005. Organic compatible solutes of halotolerant and halophilic microorganisms. *Saline. Systems.* **1**:5.

Robertson, D.E., Noll, D., and Roberts, M.F. 1992. Free amino-acid dynamics in marine methanogens - beta-amino acids as compatible solutes. *J. Biol. Chem.* **267**:14893-14901.

Robin, S., Daudin, J.J., Richard, H., Sagot, M.E., and Schbath, S. 2002. Occurrence probability of structured motifs in random sequences. *Journal of Computational Biology* **9**:761-773.

Rodriguez, R., Chinea, G., Lopez, N., Pons, T., and Vriend, G. 1998. Homology modeling, model and software evaluation: three related resources. *Bioinformatics.* **14**:523-528.

Rose, G.D., Fleming, P.J., Banavar, J.R., and Maritan, A. 2006. A backbone-based theory of protein folding. *Proc. Natl. Acad. Sci. U. S. A.* **103**:16623-16633.

Santoro, M.M. and Bolen, D.W. 1988. Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants. *Biochemistry* **27**:8063-8068.

Santos, H. and da Costa, M.S. 2001. Organic solutes from thermophiles and hyperthermophiles. *Methods Enzymol.* **334**:302-315.

Santos, H. and da Costa, M.S. 2002. Compatible solutes of organisms that live in hot saline environments. *Environ. Microbiol.* **4**:501-509.

Santos, H., Lamosa, P., Burke, A. and Maycock, C. 1999. Thermostabilization, osmoprotection, and protection against diseccation of enzymes, cell components and cells by di-glycerol-phosphate. Patent nº **98670002.9**.

Santos, H., Lamosa, P., Faria, T.Q., Borges, N., and Neves, C. 2007. The physiological role, biosynthesis and mode of action of compatible solutes from (hyper)thermophiles. In *Physiology and biochemistry of extremophiles*, (eds. C. Gerday and N. Glandorf), pp 86-103, CRC Press.

Santos, H., Lamosa, P., Borges, N., Gonçalves, L.G., Pais, T.M., and Rodrigues, M.V. 2011. Organic compatible solutes of prokaryotes that thrive in hot environments: the importance of ionic compounds for thermostabilization. In *Extremophiles Handbook*, (eds. K. Horikoshi, G. Antranikian, A. T. Bull, F. T. Robb, K. O. Stetter), pp 497-520, Springer, Tokyo.

Santos, H., Ramos, A., and da Costa, M.S. 1998. Thermostabilization, osmoprotection, and protection against desiccation of enzymes, cell components, and cells by mannosylglycerate. Patent nº **97670002.1-2105**.

Schellman, J.A. 1993. The relation between the free-energy of interaction and binding. *Biophysical Chemistry* **45**:273-279.

Schellman, J.A. 1987a. The thermodynamic stability of proteins. *Annu. Rev. Biophys. Biophys. Chem.* **16**:115-137.

Schellman, J.A. 1987b. Selective binding and solvent denaturation. *Biopolymers* **26**:549-559.

Schellman, J.A. and Schellman, C.G. 1997. Kaj Ulrik Linderstrøm-Lang (1896-1959). *Protein Sci.* **6**:1092-1100.

Schwartz, A.L. and Cutnell, J.D. 1983. One- and two-dimensional NMR studies of exchanging amide protons in glutathione. *J. Magn Reson.* **53**:398-411.

Serrano, L., Kellis, J.T., Cann, P., Matouschek, A., and Fersht, A.R. 1992. The folding of an enzyme. 2. Substructure of barnase and the contribution of different interactions to protein stability. *J. Mol. Biol.* **224**:783-804.

Shortle, D. and Meeker, A.K. 1989. Residual structure in large fragments of staphylococcal nuclease: effects of amino acid substitutions. *Biochemistry* **28**:936-944.

Shortle, D.R. 1996. Structural analysis of non-native states of proteins by NMR methods. *Curr. Opin. Struct. Biol.* **6**:24-30.

Sieker, L.C., Stenkamp, R.E., Jensen, L.H., Prickril, B., and Legall, J. 1986. Structure of rubredoxin from the bacterium *Desulfovibrio desulfuricans*. *Febs Letters* **208**:73-76.

Sieker, L.C., Stenkamp, R.E., and LeGall, J. 1994. Rubredoxin in crystalline state. *Methods Enzymol.* **243**:203-216.

Silva, Z., Borges, N., Martins, L.O., Wait, R., da Costa, M.S., and Santos, H. 1999. Combined effect of the growth temperature and salinity of the medium on the accumulation of compatible solutes by *Rhodothermus marinus* and *Rhodothermus obamensis*. *Extremophiles* **3**:163-172.

Skelton, N.J., Palmer, A.G., Akke, M., Kordel, J., Rance, M., and Chazin, W.J. 1993. Practical aspects of two-dimensional proton-detected ^{15}N spin relaxation measurements. *J. Mag. Res. B* **102**:253-264.

Sosnick, T.R. and Trewhella, J. 1992. Denatured states of ribonuclease A have compact dimensions and residual secondary structure. *Biochemistry* **31**:8329-8335.

Sugase, K., Lansing, J.C., Dyson, H.J., and Wright, P.E. 2007. Tailoring relaxation dispersion experiments for fast-associating protein complexes. *J. Am. Chem. Soc.* **129**:13406-13407.

Szilagyi, A. and Zavodszky, P. 2000. Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comprehensive survey. *Structure.* **8**:493-504.

Tadeo, X., Castano, D., and Millet, O. 2007. Anion modulation of the $^{1}H/^{2}H$ exchange rates in backbone amide protons monitored by NMR spectroscopy. *Protein Sci.* **16**:2733-2740.

Tadeo, X., Lopez-Mendez, B., Castano, D., Trigueros, T., and Millet, O. 2009. Protein stabilization and the Hofmeister effect: The role of hydrophobic solvation. *Biophysical Journal* **97**:2595-2603.

Tanford, C. 1968. Protein denaturation. *Adv. Protein Chem.* **23**:121-282.

Tanford, C. 1970. Protein denaturation. C. Theoretical models for the mechanism of denaturation. *Adv. Protein Chem.* **24**:1-95.

Tang, K.E. and Dill, K.A. 1998. Native protein fluctuations: the conformational-motion temperature and the inverse correlation of protein flexibility with protein stability. *J. Biomol. Struct. Dyn.* **16**:397-411.

Timasheff, S.N. 2002. Protein-solvent preferential interactions, protein hydration, and the modulation of biochemical reactions by solvent components. *Proc. Natl. Acad. Sci. U. S. A.* **99**:9721-9726.

Timasheff, S.N. 1992. Water as ligand: preferential binding and exclusion of denaturants in protein unfolding. *Biochemistry* **31**:9857-9864.

Timasheff, S.N. 1993. The control of protein stability and association by weak interactions with water: how do solvents affect these processes? *Annu. Rev. Biophys. Biomol. Struct.* **22**:67-97.

Tjandra, N., Kuboniwa, H., Ren, H., and Bax, A. 1995. Rotational dynamics of calcium-free calmodulin studied by ^{15}N-NMR relaxation measurements. *Eur. J. Biochem* **230**:1014-1024.

Tjandra, N., Wingfield, P., Stahl, S., and Bax, A. 1996. Anisotropic rotational diffusion of perdeuterated HIV protease from ^{15}N NMR relaxation measurements at two magnetic fields. *J. Biomol. NMR* **8**:273-284.

Tollinger, M., Skrynnikov, N.R., Mulder, F.A., Forman-Kay, J.D., and Kay, L.E. 2001. Slow dynamics in folded and unfolded states of an SH3 domain. *J. Am. Chem. Soc.* **123**:11341-11352.

Tsai, A.M., Udovic, T.J., and Neumann, D.A. 2001a. The inverse relationship between protein dynamics and thermal stability. *Biophysical Journal* **81**:2339-2343.

Tsai, C.J., Kumar, S., Ma, B.Y., and Nussinov, R. 1999. Folding funnels, binding funnels, and protein function. *Protein Science* **8**:1181-1190.

Tsai, C.J., Ma, B.Y., Sham, Y.Y., Kumar, S., and Nussinov, R. 2001b. Structured disorder and conformational selection. *Proteins* **44**:418-427.

Tucker, P.W., Hazen, E.E. Jr., and Cotton, F.A. 1978. Staphylococcal nuclease reviewed: a prototypic study in contemporary enzymology. I. Isolation; physical and enzymatic properties. *Mol. Cell Biochem.* **22**:67-77.

Turner, D.L., Brennan, L., Meyer, H.E., Lohaus, C., Siethoff, C., Costa, H.S., Gonzalez, B., Santos, H., and Suarez, J.E. 1999. Solution structure of plantaricin C, a novel lantibiotic. *European Journal of Biochemistry* **264**:833-839.

Turoverov, K.K., Kuznetsova, I.M., and Uversky, V.N. 2010. The protein kingdom extended: ordered and intrinsically disordered proteins, their folding, supramolecular complex formation, and aggregation. *Prog. Biophys. Mol. Biol.* **102**:73-84.

Umetsu, M., Tsumoto, K., Hara, M., Ashish, K., Goda, S., Adschiri, T., and Kumagai, I. 2003. How additives influence the refolding of immunoglobulin-folded proteins in a stepwise dialysis system-spectroscopic evidence for highly efficient refolding of a single-chain FV fragment. *J. Biol. Chem.* **278**:8979-8987.

Uversky, V.N., Oldfield, C.J., and Dunker, A.K. 2008. Intrinsically disordered proteins in human diseases: introducing the D2 concept. *Annu. Rev. Biophys.* **37**:215-246.

Varley, P.G. and Pain, R.H. 1991. Relation between stability, dynamics and enzyme activity in 3-phosphoglycerate kinases from yeast and *Thermus thermophilus*. *J. Mol. Biol.* **220**:531-538.

Vedadi, M., Niesen, F.H., lali-Hassani, A., Fedorov, O.Y., Finerty, P.J., Wasney, G.A., Yeung, R., Arrowsmith, C., Ball, L.J., Berglund, H., Hui, R., Marsden, B.D., Nordlund, P., Sundstrom, M., Weigelt, J., and Edwards, A.M. 2006. Chemical screening methods to identify ligands that promote protein stability, protein crystallization, and structure determination. *Proc. Natl. Acad. Sci. U. S. A.* **103**:15835-15840.

Vidugiris, G.J.A., Markley, J.L., and Royer, C.A. 1995. Evidence for a molten globule-like transition-state in protein folding from determination of activation volumes. *Biochemistry* **34**:4909-4912.

Vihinen, M. 1987. Relationship of protein flexibility to thermostability. *Protein Eng.* **1**:477-480.

Wang, A., Robertson, A.D., and Bolen, D.W. 1995. Effects of a naturally occurring compatible osmolyte on the internal dynamics of ribonuclease A. *Biochemistry* **34**:15096-15104.

Wang, J.F., Hinck, A.P., Loh, S.N., LeMaster, D.M., and Markley, J.L. 1992a. Solution studies of staphylococcal nuclease H124L. 2. ^{1}H, ^{13}C, and ^{15}N chemical shift assignments for the unligated enzyme and analysis of chemical shift changes that accompany formation of the nuclease-thymidine 3',5'-bisphosphate-calcium ternary complex. *Biochemistry* **31**:921-936.

Wang, J.F., Mooberry, E.S., Walkenhorst, W.F., and Markley, J.L. 1992b. Solution studies of staphylococcal nuclease H124L. 1. Backbone ^{1}H and ^{15}N resonances and secondary structure of the unligated enzyme as identified by three-dimensional NMR spectroscopy. *Biochemistry* **31**:911-920.

Watson, E., Matousek, W.M., Irimies, E.L., and Alexandrescu, A.T. 2007. Partially folded states of staphylococcal nuclease highlight the conserved structural hierarchy of OB-fold proteins. *Biochemistry* **46**:9484-9494.

Woessner, D.E. 1962. Spin relaxation processes in a two-proton system undergoing anisotropic reorientation. *Journal of Chemical Physics* **36**:1-4.

Wooll, J.O., Wrabl, J.O., and Hilser, V.J. 2000. Ensemble modulation as an origin of denaturant-independent hydrogen exchange in proteins. *J. Mol. Biol.* **301**:247-256.

Wright, P.E. and Dyson, H.J. 2009. Linking folding and binding. *Current Opinion in Structural Biology* **19**:31-38.

Yang, D.W. and Kay, L.E. 1996. Contributions to conformational entropy arising from bond vector fluctuations measured from NMR-derived order parameters: Application to protein folding. *J. Mol. Biol.* **263**:369-382.

Yip, K.S., Stillman, T.J., Britton, K.L., Artymiuk, P.J., Baker, P.J., Sedelnikova, S.E., Engel, P.C., Pasquo, A., Chiaraluce, R., Consalvi, V., Scandurra, R., and Rice, D.W. 1995. The structure of *Pyrococcus furiosus* glutamate dehydrogenase reveals a key role for ion-pair networks in maintaining enzyme stability at extreme temperatures. *Structure.* **3**:1147-1158.

Zartler, E.R., Jenney, F.E., Terrell, M., Eidsness, M.K., Adams, M.W.W., and Prestegard, J.H. 2001. Structural basis for thermostability in aporubredoxins from *Pyrococcus furiosus* and *Clostridium pasteurianum*. *Biochemistry* **40**:7279-7290.

Zavodszky, P., Kardos, J., Svingor, A and Petsko, G.A. 1998. Adjustment of conformational flexibility is a key event in the thermal adaptation of proteins. *Proc. Natl. Acad. Sci. U. S. A.* **95**:7406-7411.

Zhang, O., Kay, L.E., Shortle, D., and Forman-Kay, J .D. 1997. Comprehensive NOE characterization of a partially folded large fragment of staphylococcal nuclease Delta131Delta, using NMR methods with improved resolution. *J. Mol. Biol.* **272**:9-20.

Zurn, A., Klenk, C., Zabel, U., Reiner, S., Lohse, M.J., and Hoffmann, C. 2010. Site-specific, orthogonal labeling of proteins in intact cells with two small biarsenical fluorophores. *Bioconjug. Chem.* **21**:853-859.