

Structural Bioinformatics: Methods, Concepts and Applications to Blood Coagulation Proteins

Bruno O. Villoutreix*

INSERM U428, University Paris V, 4 Ave. de L'Observatoire 75006 Paris, France



Abstract: Structural and theoretical analyses of proteins are central to the understanding of complex molecular mechanisms and are fundamental to the drug discovery process. Computational techniques yield useful insights into an ever-wider range of biomolecular systems. Protein three-dimensional structures and molecular functions can be predicted in some circumstances, while experimental structures can be analyzed in depth via such computational approaches. Non-covalent binding of biomolecules can be understood by considering structural, thermodynamic and kinetic issues, and theoretical simulations of such events can be attempted. The central role of electrostatic interactions with regard to protein function, structure and stability has been investigated and some electrostatic properties can be modeled theoretically. Computer methods thus help to prioritize, design, analyze and rationalize biochemical experiments. Cardiovascular diseases and associated blood coagulation disorders are leading causes of death worldwide. Blood coagulation involves more than 30 proteins that interact specifically with various degrees of affinity. Many of these molecules can also bind transiently to phospholipid surfaces. Numerous point mutations in the genes of coagulation proteins and regulators have been identified. Understanding the coagulation cascade, its regulation and the impact of mutations is required for the development of new therapies and diagnostic tools. In this review, we describe concepts and methods pertaining to the field of structural bioinformatics. We provide examples of applications of these approaches to blood coagulation proteins and show that such studies can give insights about molecular mechanisms contributing to cardiovascular disease susceptibility.

1. INTRODUCTION

Nature has evolved over millions of year highly sophisticated mechanisms to perform biological functions. Subtle imbalances in a living organism or apparently insignificant errors during some chemical processes can lead to disease states. Proteins carry out numerous functions in the human body, and studies investigating the relationship between sequence-structure-function and energetics aim at, fundamental understanding of molecular mechanisms and of protein folding and at the development of new and more specific therapeutic compounds.

Knowledge of protein three-dimensional (3D) structures is of major importance in providing insights into their molecular functions [1, 2]. Analysis of 3D structures assists identification of binding sites and thus facilitates design of new drugs. Also, structural investigations at atomic resolution help in conceiving how a single point mutation in a gene, and the possible subsequent amino acid substitution at the protein level can, for instance, lead to protein deficiency and disease. Moreover, biochemical methods to probe protein functions, or to enhance/inhibit some reactions, and the entire protein engineering field benefit tremendously when scientists have access to structural information prior to initiating research projects. Such data

indeed tend to simplify the design of experiments and the rational processing of experimental results [3-5]. Clearly, sophisticated methods used to compare genes and/or amino acid sequences can also provide significant amount of information about molecular functions, whether applied alone or in the context of a three-dimensional representative of a family member and facilitate rational engineering [6-11].

X-ray crystallography, NMR and biochemical methods have been essential to start developing concepts about how amino acid sequences relate to folding and function [12]. Such experimental data have helped in the design of computer methods facilitating prediction and analysis of protein structures [13-28]. As new 3D folds [29] are reported (e.g., via the structural genomics initiative) at the Protein Data Bank [30] (<http://www.rcsb.org/pdb/>), our level of understanding is increasing because structural data provide more powerful links to function than does sequence information alone [31, 32]. Along the same line of reasoning, reports about experimentally determined protein-protein complexes should help understanding molecular interactions and a Database of Interacting Proteins has been created [33]. Furthermore, completion of large-scale genome sequencing projects has resulted in the discovery of thousands of new protein sequences that will also help in relating sequence, fold and functions [6, 34-36]. However, the vast majority of these new proteins has not been studied by crystallography or NMR, nor they will be for many years to come [17]. Furthermore, it might be an impossible task to study some proteins or protein complexes by NMR or crystallography alone. Therefore, theoretical techniques

*Address correspondence to this author at the INSERM U428, University Paris V, 4 Ave. de L'Observatoire, 75006 Paris, France; Email: villoutreix@pharmacie.univ-paris5.fr

allowing prediction of 3D structures, of molecular complexes and of functions have been developed over the years and significant efforts are still expended in improving these computational methods.

Bioinformatics encompasses tools and concepts aiming at the analysis of gene expression (e.g., cDNA microarrays immobilized on slides or gene chips) as well as approaches leading to the prediction of protein structures and functions. The term structural bioinformatics has been coined recently, and commonly refers to research studies dealing with structural predictions, simulations, sequence and structure analyses, family classification, and genome annotation. The field of structural bioinformatics and related concepts provide not only a means of organizing one's thinking about sequence-structure-function problems, but also a framework for predicting unobserved behavior and suggesting novel experiments.

In the first part of this review, we comment on computational methods and key concepts pertaining to the field of structural bioinformatics. In the second part, we present illustrative applications of these methods to coagulation proteins. This review is far from exhaustive, as the topics under examination are far too vast to be summarized in a single article. We therefore reference several review papers as well as research articles that can guide readers to in-depth description of the material under consideration or that underline some fundamental notions with specific case studies in various areas of Life Science research.

2. STRUCTURAL BIOINFORMATICS: METHODS AND CONCEPTS

Bioinformatics has been referred to as an interdisciplinary science that involves both conceptual and practical tools for the understanding, generation, processing and propagation of biological information [37]. There are currently three main trends in the field of structural bioinformatics: protein folding and 3D structure predictions (e.g., development of more accurate free energy functions), investigation of intermolecular interactions and of macromolecular complexes (possible applications are structure-based ligand design, scoring functions for *in silico* screening, metabolic pathways), and prediction of potential structural impacts of amino acid substitutions on protein folding and stability (e.g., conformational diseases). Studies on this latter topic have been stimulated by the realization that misfolding and/or aggregation are involved in a number of diseases (e.g., naturally occurring mutations can affect protein folding and can lead to disease states, such diseases are often referred to in the literature as conformational diseases).

Numerous debates are taking place in the field of structural bioinformatics, for instance about the stabilizing/destabilizing effects of ionic interactions, either within one protein or at the interface between two molecules. Different opinions as to the nature of ligand binding surfaces (hydrophobic, polar, charged) can also be noted in the literature. With regard to possible roles of amino acid substitutions on protein folding and stability, several

hypotheses have been put forward. Assuming a simple two-state folding process, with a protein moving from the unfolded state through a transition state to the folded state, thermodynamic stability depends on the relationship between the Gibbs free energies of the folded and unfolded states (see for instance [38]). Mutations will affect both the folded and the unfolded states. A key problem here is the lack of structural knowledge about the unfolded state. Also, water molecules can play very important roles, not only in the stabilization or destabilization of molecules, but also at the interface between two molecules for appropriate docking and binding, or they can compensate for mutations at protein-protein interfaces. Water molecules can also contribute to the coordination of ions. However, the positions of these water molecules are difficult to define experimentally and theoretically.

All methods have strengths and weaknesses, and with regard to structure determination via NMR or crystallography, 3D conformation of molecules may be under the influence of crystal packing effects and experimental conditions (pH, protein concentration, etc.). For instance, if the energy barrier between two conformations is low, different conformations can be found under different experimental conditions [39]. In a similar vein, often only one conformation is available for structural analysis, and it is difficult to know if this structure represents the true *in vivo* conformation or results from a given set of experimental conditions. Also, some regions are flexible, and side chains can re-orient. Furthermore, rules and programs that apply to soluble proteins are not always appropriate to study membrane proteins. In the following paragraphs, we discuss and comment on some of the points mentioned above with an emphasis on protein structural predictions, molecular simulation, electrostatics, docking and conformational diseases (also named biostructural pathology).

2.1. Structural Prediction and Comparative Model Building

At present, techniques to predict protein structures can be divided in two groups, *ab initio* and comparative model building methods [40-45]. *Ab initio* methods attempt to fold a sequence in 3D using only information about the sequence and mathematical models of interactions between amino acids, while comparative modeling requires the presence of experimental templates. The complexity of conformational space has so far precluded *ab initio* prediction of 3D structures of large proteins, whereas comparative (or homology) modeling is a much more reliable approach.

The field of *ab initio* protein folding is progressing, and special potential energy functions for folding simulations are under development [46-48]. Lattice models [49-51] and genetic algorithms associated with some modified energy functions have been used for protein structure predictions [52]. Also, it is encouraging to note that some peptide structures could be predicted using well established methods such as molecular dynamics or Monte Carlo searches [53-56]. Simplified representations of amino acid residues, together with some filtering functions, have proven a successful approach for predicting the structure of short

peptides [57]. Experimental and *ab initio* 3D model structures (and others) can be refined via molecular dynamics (see for instance [58-60] and references therein). The possibility of predicting disulfide bonds would be very valuable to all structural prediction methods including *ab initio* approaches. It is feasible to predict the oxidation state of cysteines in proteins ([61, 62]) but predicting the pairs of cysteines that form disulfide bonds in proteins has not been achieved so far.

It is known that 3D folds have evolved at a much slower rate than have amino acid sequences, and that proteins sufficiently similar at the amino acid sequence level (having even as low as 5 % sequence identity) can have similar 3D structures [63]. Comparative modeling is a method to build 3D models for proteins for which crystal or NMR structures (called templates) of several family members are known. It is well recognized that predicted models can serve as an excellent basis for identifying binding sites, for protein engineering, and drug design [45, 64-66]. Yet, different levels of structural accuracy can be expected when developing a model [67-69]. For instance, a 3D model is more difficult to develop if several large insertions and/or deletions between the protein to build and the experimental templates are present. Moreover, a theoretical model can only be accurate if the sequence alignment used during the modeling procedure is reliable [70] (in other words, energy refinement protocols usually can not easily fix sequence misalignments). To circumvent in part this problem, several sequence analysis methods can be utilized during comparative model building experiments. Different energy functions can also be used during the refinement procedures. In addition, several sequence to structure alignments can be tested.

Information about sequence alignment methods such as pairwise and multiple sequence alignments, domain parsing, family and phylogenetic classifications, prediction of signal peptides, identification of glycosylation sites, and secondary structure predictions can be found in a recent review by Xu and collaborators [71]. Similarly, numerous tools and Web addresses, as well as information about threading, homology modeling, energy refinement and docking are presented in their review and will therefore not be mentioned here. New computer tools helping to build protein models and to understand molecular functions will however be listed in the different sections of the present review.

In general, a 3D structure can be predicted accurately by comparative modeling when the sequence identity between the template(s) and the model to build is high (50-60% and above) with few insertions and deletions. When the identity between the sequence of the protein to be predicted and the one of the experimental 3D template is low (25-30 %), the model tend to be less accurate in some regions. However, the accuracy of the models increases if several 3D templates are available, such that it becomes possible to generate precise sequence to structure alignments. The predicted structures are in general less accurate when sequence identity between the sequence of the protein to model and the template is very low (e.g., after threading, the sequence identity between the model to build and the template can be around 10 %). This problem can be partially circumvented if multiple sequence

alignments [72] and structural alignments are available [73-76]. In general, when the sequence identity between the protein to build and the templates is high, fully automated structure prediction methods can be used. However, it is clear in most situations that "human plus machine" predictions are superior to automated ones [77].

Numerous studies have shown that proteins with seemingly (using traditional computer alignment programs) unrelated sequences could adopt similar folds. Within this context, conventional sequence alignment procedures can be "lost" when the sequence identity falls below 20 %. Empirical observation of sequences and 3D structures led to the concept of "twilight zone". This represents a range of sequence identity that sets boundary confidence levels for detecting evolutionary relatedness among proteins in conventional sequence alignment analysis [78]. Attempts to understand and define the twilight zone in sequence-structure space are of course important in comparative modeling projects [78]. Because of all these observations (i.e., structures are more conserved than amino acid sequences), the notion of fold recognition has emerged. In protein fold recognition or threading, modelers search to find if the sequence of a protein of unknown structure is compatible with the fold of proteins having known 3D structures [43, 73, 79-81]. Because the approach is different from the classical sequence to sequence comparison methods, threading algorithms can be very efficient in identifying structural templates when the sequence identity is very low [82-84]. In practice, once a structural template has been identified using fold recognition, multiple sequence and structural alignments have to be carried out. The other steps to build a model follow those used in comparative modeling. In analyzing the sequence of a new protein with no known function, having an unknown 3D structure, and apparently not belonging to any family, threading can at once help to identify remote homologues and predict structure and (molecular) functions. New methods to refine model structures resulting from comparative modeling and threading are under development [85].

The modeling of loops (or of some small insertion domains) can be very challenging. For example, loop structures can be flexible and might not be visible via NMR or X-ray experiments. In comparative modeling, problems with predictions of large insertion regions can be due to a lack of templates or to inappropriate forcefield parameters and energy functions. Thus, loop structures or insertions should always be analyzed with caution. Several tools are, however, available to try to circumvent this problem. Simulation techniques of fully solvated molecules tend to be computationally prohibitive and alternative approaches are required. Among some recently reported methods, Monte Carlo-Molecular Dynamics sampling using a generalized Born model of solvation effects seem promising [86]. A database of about 8000 loops of up to 20 residues in length has been recently created [87], and this represents a "knowledge-based" alternative for loop building (i.e., use of known protein 3D fragments). Some methods combine a knowledge-based approach and *ab initio* (or conformational searching) algorithms [88]. The modified energy function reported by Sali and colleagues [89] is another very interesting approach. In this latter study, it was noted that

the average accuracy of prediction was limited primarily by the accuracy of the energy function rather than by the extent of conformational sampling, suggesting that new version of such energy function could lead in the future to accurate loop prediction protocols. Possibly, new methods combining *ab initio* and docking approaches could be developed to build missing insertion domains in theoretical or experimental structures.

The reliability of protein models can be evaluated not only via interactive structural analysis [65] but also using theoretical means such as, effective energy functions [27, 90, 91], 3D profile [92] and statistical potentials [93, 94]. These types of methods can, in some circumstances, be sufficiently precise to discriminate well-folded from misfolded protein models.

2.2. Flexibility, Plasticity and Domain Motions

Proteins are not static but some (important) degrees of flexibility have been noticed when analyzing experimental structures or when trying to match fluorescence resonance energy transfer or small angle X-ray scattering data with experimental 3D structures. Movements can take place at the surface (e.g., loops, side chains) but also in the core interior (core plasticity, e.g., proteins can be regarded in some ways as liquid-like although by some criteria they look-like organic crystals)[95], where molecules can undergo varying degrees of conformational changes, especially during interactions with ligands. Different types of movements can take place and for example involve entire segments. Further, allosteric behavior, either discussed in term of effectors inducing a new structure or as inducing a dynamic population shift (and stabilization of one conformation) in preexisting equilibrium, implies flexibility. Protein folding reactions are dynamics in nature while hinge bending and domain motions also reflect flexibility. Such movements can be crucial to the function of a protein. In fact, because the 3D structures of proteins change rapidly over time, a more realistic representation of 3D structures could be dynamic trajectories. Method like molecular dynamics or related approaches allow theoretical investigations of some of these conformational changes [60, 96-98]. One important problem with such approaches is that if proteins are simulated in their fully solvated state, eventually using periodic boundary conditions, 90% of the calculations is spent on moving water molecules and on others relatively uninteresting high frequency motions. Another problem is the length of the simulations. Long simulations could be performed [99], however, it was found that individual trajectories of length up to 5 ns sample only a fraction of the conformational distribution generated by ten independent 120 ps trajectories at 300 K [100]. Electrostatic interactions in such calculations are also difficult to handle, even if an approach (with numerous limitations, see for instance [101, 102] and references therein) for simulation of molecules with explicit solvent under periodic boundary conditions, like the particle mesh Ewald algorithm, has been proposed [103, 104].

Important efforts are presently dedicated to the development and assessment of "old"-new methods treating solvent molecules implicitly [105, 106]. Generalized Born

model to replace solvent seems promising [107-109]. Further, method like multi-body order dynamics in which the protein is sub-structured into flexible and rigid bodies has been proposed to attempt to simulate large motions relatively fast [110, 111]. While interesting, it is not always straightforward to decide which part of a protein belongs to which structural subgroups. The Gaussian network model has also been used to simulate large-scale motions [112]. Also, the method of principal component analysis has been applied to molecular dynamics trajectories to study protein motions [113, 114] (see also the paper by Amadei *et al.* [115] to filter out high frequency motions).

Molecular dynamics can be used before designing experiments to evaluate potential structural impacts of mutations (e.g., destabilization) [116-120]. For such studies, simulations can be performed at room temperature, higher temperatures or via simulated annealing protocols. As mentioned earlier, molecular dynamics is often used to search conformational space. Running dynamics at higher temperatures, on the order of 400 to 600 K, gives the molecule enough energy to overcome relatively large energy barriers and in this way the conformational space is searched more efficiently while the process also speeds-up the calculation (such calculations have to be run with care). Simulated annealing is a special case of simulation in which the temperature is gradually reduced during the calculations. This method is often used to investigate mutated structure or to probe flexible areas.

Simulations of protein folding and unfolding and overall comparisons with experimental data have also been reported [121, 122]. Unfolded proteins may not be fully extended but it seems indeed that unfolded states represent an ensemble of poorly structured conformers that yet can retain significant residual structures and native-like contacts. Considering the importance of conformational changes, inherently dynamics in nature, simulations can provide new insights not only to understand the mechanisms involved in for instance conformational diseases but also to develop a new generation of rationally devised chemotherapeutically active reagents [123-125]. Method such as graph theory (that analyzes bond networks in proteins and identify flexible and rigid regions a million times faster than molecular dynamics) can be applied to proteins [126].

2.3. Electrostatics

Electrostatic interactions such as salt bridges and their networks play important roles in protein folding (and denatured states), stability, binding (and other molecular functions) and structure [127-132]. For example, electrostatic plays a crucial role in enzyme catalysis, in part due to a pre-organized polar environment within the active site [133, 134]. Also, electrostatic calculations may help identify residues that are potentially important for functions (the hypothesis being that functional residues often destabilize proteins) (see [127] and references therein). Many simulation models have been used to study protein electrostatics. These range from microscopic models (e.g., density functional calculations, hybrid quantum mechanics-molecular mechanics, molecular dynamics or Monte Carlo

simulations), semi-macroscopic models (e.g., dipolar models) to macroscopic models (e.g., Poisson-Boltzmann, modified Tanford-Kirkwood model, generalized Born model, screened Coulomb potentials) [102, 135-145]. Calculations of pKa values can be performed but are not trivial [102, 146-159]. For instance, within the Poisson-Boltzmann scheme, it is important but difficult to decide which dielectric constant should be assigned to the protein interior and to define the meaning of dielectric constant [102]. Calculations using the Fröhlich-Kirkwood dipole moment fluctuation model evaluate the dielectric constant of proteins to range from 4 to 15 to 40 (see meanings of dielectric constant in [102, 160, 161]). High apparent dielectric constants in the interior of proteins could reflect water penetration [162] (see also notion of packing density [95]). Yet, further enhancements of the nonlinear Poisson-Boltzmann equation with multiple dielectric constants and new method to decompose the energy terms to remove dependency on the lattice grid (into which the proteins are mapped) have been reported [163]. Calculations of pH-dependent properties can be accomplished using the modified Tanford-Kirkwood theory [164] or "modified" Poisson-Boltzmann equation [147, 152] and information about charges of titratable groups at a given pH can be used as input for Poisson-Boltzmann calculations [165-168]. Electrostatic computations have also been used to map the free energy of water molecules around a protein in order to facilitate understanding of protein-water interactions [169]. Principles for enhancing the effect of electrostatic interactions in molecular binding have been discussed [170]. Structural alignments based upon electrostatic potentials can provide important information about protein functions [171]. Electrostatic potentials can be visualized either by displaying isopotential surfaces or by color-coding the molecular surface according to potential values (e.g., red for negative potentials and blue for positive potentials while a linear interpolation is used to color surface potentials between the two selected values). Electrostatic potential visualization can definitively help in the identification of functional hot spots at the surface of a protein [135].

Information about the stabilizing and destabilizing characters of ionic interactions can be computed and investigated experimentally. The roles of ionic interactions with regard to protein stability are still under debate [102]. Using continuum electrostatic calculations, it has been found that salt-bridges were generally destabilizing with respect to mutation to hydrophobic isosteres (often the computed data are pH-independent since ionizable groups are assumed to be in their charged states and calculations are performed at pH = 7) [172]. Some experimental evidences support such data, since for example, contribution of buried salt-bridges (peptide in octanol environment to mimic protein interior) to stability was found to be neutral or slightly destabilizing and in any case were less stabilizing than hydrophobic residues [173]. However, more recently, theoretical analysis of ionic interactions suggested most salt-bridges to be stabilizing [174, 175]. These latter calculations are also consistent with many experimental data [38]. Three factors influence the energetics of salt-bridges: the geometry, the location and the interaction of salt-bridging residues with other charged residues in the protein. Optimal salt-bridge geometry and appropriate interactions with other charges can compensate desolvation penalty. Buried salt-bridges could thus be

strongly stabilizing (e.g., stabilize proteins by 3-5 kcal/mol) [149] and would tend to be more stabilizing than surface-exposed ones (e.g., stabilize proteins by 0.5 kcal/mol). Overall, when salt-bridges are part of a network, they are most likely stabilizing. Salt-bridges and their networks may easily break and reform in solution [176]. Possibly, ionic interactions can oppose binding in some cases [177] but continuum electrostatic calculations can be very sensitive to the presence of water molecules. Thus, for the time being, the impact of ion-pairs on protein stability needs to be considered with care [102]. Quantitative calculations are not trivial as assumptions about the unfolded structure, dielectric constant and ionization states, among other parameters, have to be made but their effects are difficult to estimate. Furthermore, protein engineering experiments usually probe several physical contributions to stability at the same time (electrostatic, steric, entropic) [149]. Overall, it is generally accepted that salt-bridge and hydrogen bond networks at the interface between two proteins or within a protein, tend to be stabilizing. Most likely, salt-bridges contribute to the specificity of the fold. Along the same line of reasoning, it is interesting to note that clusters of similar charge could also be stabilizing. For instance, two arginines close in space can stabilize a protein through interaction with many water molecules [178].

Although not always applicable to mesophilic proteins, structural analysis of hyperthermophilic proteins, halophilic (adapted to high-salt conditions) proteins and psychrophilic (cold-adapted) proteins gives important insights about specific features increasing stability under extreme environments. Considering only electrostatic properties, hyperthermophilic proteins can achieve enhanced stability through buried ion-pair network within inter-subunit regions [179] or via increasing the number of ion-pairs [38, 180, 181]. This would be consistent with the fact that at high temperature, the hydrophobic effect would be minimal (i.e., it seems easier for Nature to improve stability via electrostatic means than through manipulating the location and the number of hydrophobic residues within the amino acid sequence). Also, at high temperatures, the desolvation penalty for formation of salt-bridges should be markedly reduced in magnitude [182]. Yet, it might not be the number of ion-pairs but rather the location of these groups within the protein structure that is essential for stability [183]. Halophilic proteins could be stabilized through displaying many acidic residues at the surface that would avoid self-aggregation [184] and possibly by better water-binding capacity as compared to their non-halophilic homologues [185]. Psychrophilic proteins could be stabilized because they possess, among other structural features, an increased number of intra-molecular ion-pairs [186, 187].

2.4. Protein-Ligand Interactions and Docking

Proteins function through their binding properties. The target molecules vary in size and complexity, from a single ion to large multi-domain proteins to macromolecular complexes. In a sense, there are many similarities (with some important differences) between binding and folding although hydrophilic side chains seem to play a more important role in binding than in folding [188-191]. Burial

of solvent-accessible surface areas is an important factor contributing to the free energy of both, folding and binding.

Protein-ligand binding affinity is determined by the Gibbs energy [$K_a = \exp(-G/RT)$] and G is given by $G = H - T S$ (where H is the change in enthalpy and S the change in entropy and K_a the equilibrium association constant). High affinity can be accomplished via negative

H (e.g., formation of polar interactions upon binding is favorable while burial of hydrophobic groups is not) and/or via positive S (e.g., desolvation of nonpolar groups upon binding, also called solvation entropy, is favorable, while loss in conformational degrees of freedom of both receptor and ligand upon binding or conformational entropy change is not). Electrostatic attraction between two molecules is enthalpic in nature, and loss of conformational entropy is often offset by massive dehydration of interfaces. Enthalpy can drive the interaction between two molecules with a negligible change in entropy on binding [192]. For instance, the rotational, translational and conformational entropy loss upon binding can be balanced by the hydrophobic effect (i.e., the increase in the entropy of the solvent due to burial of surface area) and by changes in the protonation states, as well as counter ion release, and the presence of new vibrational degrees of freedom [193-195]. On the other hand, a favorable entropy change can be the driving force for ligand binding [196]. Thus, strong binding would require that the ligand be sufficiently hydrophobic, rigid (both hydrophobicity and rigidity can affect entropy) and polar (polarity strongly influences enthalpy). Protein interfaces are close-packed, and several types of direct interactions have been noticed, such as hydrogen bonds, van der Waals, and ionic interactions, together with contributions of solvent entropy. Hydrophobicity is a major factor that stabilizes protein-protein associations, while polar/electrostatic interactions and electrostatic complementarity, but not necessarily shape complementarity, contribute to specificity [197-200]. Yet, in some situations, both, specificity and affinity can arise mainly from shape complementarity [201]. It is also important to differentiate the notion of charge complementarity from the notion of electrostatic complementarity. For instance, there might be significant electrostatic complementarity at protein-protein interfaces but little charge complementarity [202]. The former can also contribute importantly to specificity. Usually, the interface between two proteins buries about 600-1000 Å² of accessible surface area with about 10 hydrogen bonds/salt bridges. Often, the two protein partners can be considered as behaving as rigid bodies, with some side chain/main chain rearrangements [203]. Contributions of the main-chain/side-chain components to the overall interaction energy can vary significantly between different types of complexes (e.g., protease-inhibitors, antibody-antigen) [204]. A single binding site can bind very different molecular structures, and such a convergent binding surface tends to be highly accessible, adaptive and hydrophobic while containing relatively few sites for polar interactions [205].

The interaction between two molecules A and B can be conceived as:

$A + B \rightleftharpoons A:B \rightleftharpoons AB$; A:B being defined as a low affinity loose complex, often called an encounter complex,

and AB being the final high affinity complex. Long-range electrostatic forces can attract and then hold A and B in close vicinity within the encounter complex [206, 207]. Short-range interactions between specific residues, expulsion of water from the interface, and possible structural rearrangements are required for the formation of the final complex. Electrostatic and/or hydrophobic forces can steer a ligand into its binding site and play an important role in the rates of diffusion of substrates. In some systems, the rate of association is dictated primarily by long-range electrostatic interaction that stabilizes the encounter complex. Once the ligand is close to the receptor, ionic tethering (e.g., loop motions, channel openings upon ligand binding) can regulate or modulate substrate binding [139]. In the last step, direct salt bridges and hydrogen bonds (among others) between the ligand and the receptor can be formed. Conserved electrostatic potentials are often noted within a family of molecules (e.g., enzymes) in the binding site region (e.g., active site). It is always difficult to evaluate the relative importance of each interaction type (e.g., hydrogen bond, ionic) because their relative contributions to binding vary [208, 209]. Strong attractions between large hydrophobic objects at large distances have been observed but the importance of such mechanisms is still under evaluation with regard to protein folding and binding [210, 211].

Two molecules can interact in several different ways and it can be difficult to differentiate between crystal and biological interfaces [212]. Site directed mutagenesis could be an efficient way to probe protein-protein interactions, yet such data have to be evaluated with caution (see examples with thrombin [213]). It was noted that despite the large size of a binding interface, few residues could contribute a large fraction to the binding free energy [214]. As such, there is little correlation between buried surface area and free energy of binding. There are hot spots at interfaces that are enriched in Trp, Tyr and Arg [215]. Whereas overall, interfaces manifest a higher frequency of occurrence of hydrophobic residues [216, 217], some residues (essentially polar) are preferentially conserved (Arg, Gln, His, Asp, Pro) [218]. Hot spots can also be located within flexible regions [219]. There are differences at the interfaces between proteins that must be independently stable before association (transient) from those present in obligate complexes. The latter interfaces tend to be depleted in charged groups and to present an abundance of hydrophobic residues [220-222]. Yet, Lo Conte *et al.* did not observe that protein-protein interaction sites (for transient binding) were enriched in hydrophobic residues, in contrast with many related studies [223]. Clusters of charged residues could also be important for protein-protein interactions [224, 225]. The role of water molecules can also be significant [226].

Numerous docking procedures, scoring functions, lead optimization and virtual screening methods have been developed during the last decade (e.g., it is possible to compute in some cases approximate G values and thus an equilibrium association constant K_a , which is the reciprocal of the dissociation constant K_d) [71, 227-245]. Clearly, the selection of a method will depend on several criteria. For instance, if the binding site is not known and if the ligand is a small molecule, screening for a cavity on the receptor surface might be needed [246]. If the ligand is a protein, a

search for relatively flat and extended surfaces, possibly presenting with solvent exposed hydrophobic clusters and/or patches of polar and charged residues could be performed. In both cases, tools to predict binding sites using multiple sequence alignments with or without 3D structure information have been developed [8, 216, 247-257]. The ability to predict the strength of noncovalent binding between macromolecules from structural parameters has been a long-standing goal in computational chemistry. Yet, docking programs can not always achieve reliable prediction at present [258-260].

A possible docking protocol is to start with rigid body docking and Monte Carlo sampling. The ligand can be considered as flexible while the protein receptor can be maintained rigid initially to facilitate the search. A second step is then to allow for relative flexibility and to run a simulated annealing protocol. The last step could be to use some scoring functions to rank the different hits. Comparisons between the resulting complexes and predicted binding sites based upon sequence analysis can further confirm the validity of the docked structures. During such a process, it is important to try to evaluate the potential roles of water molecules and ions, as their presence or absence can change considerably the calculation outputs. Also, glycans are seldom present on experimental and modeled protein 3D structures, while they could play a role during binding [261]. The docking protocol described above (with attempts to derive binding energies) relates to the thermodynamic aspects of protein-protein interactions, but protein-protein association kinetics can also be important to probe, and such events can be modeled via Brownian dynamics (see [259] and references therein).

Better understanding of protein-membrane interactions and of the behavior of phospholipids allow computer simulations of such systems (see for instance [154, 262-264]). Along this line, a preference for aromatic residues at protein-membrane interfaces has been observed, and this can help the docking process [265]. Several studies dealing with experimental/theoretical analysis of membrane binding properties of blood coagulation proteins and other molecules (e.g., peptides) with a special emphasis on electrostatics have been reported [266-270].

Protein-ligand interactions can be probed using site directed mutagenesis [213]. Yet, it is important to assess that the mutation does not damage the 3D structure and stability of the molecules (the quality of the mutant 3D structure can be assessed via testing binding of several monoclonal antibodies possessing different epitopes). Double or higher-order mutant cycles can be used to analyze interactions [209]. Evaluation of binding sites using small peptides without knowledge of 3D structures or at least without performing sequence analysis can give ambiguous results. For instance it is common that a peptide inhibits an interaction while, once mapped onto the protein structure, the segment is seen fully buried into the protein core, and thus it can not form a recognition site under physiological conditions. Linear peptides tend to be flexible, and they can be forced to adopt a specific conformation when they get into contact with a receptor. Yet, short amino-acid stretches can be very helpful for rational drug design studies, and it is

possible to use computer methods to move from peptides to non-peptide peptidomimetics (see example of inhibitors against aspartic peptidases by Rich and colleagues [271] and references therein). Probing binding sites using antibodies can be helpful, but the epitope should be known and the relative size of the antibody versus the receptor should be considered [272]. Clearly, antibodies are such large molecules that they can interfere with binding because of steric hindrance and not because they cover a binding surface. The ionic versus non-ionic nature of an interaction can be tested via experiments performed at different salt concentrations [273-275], and such information can be used to facilitate evaluation of computer docking protocols.

2.5. Biostructural Pathology and Conformational Diseases

Biological systems have evolved elaborate procedures to ensure that proteins fold correctly. Despite these quality control events, a range of debilitating diseases is associated with protein misfolding and aggregation [124, 276-278]. Numerous studies attempt to understand such problems at a structural level. The term biostructural pathology has thus been used to describe these phenomena [279]. Biostructural was selected to differentiate these studies performed at atomic resolution from the well-established discipline of structural pathology. Conformational disease (e.g., protein aggregation after a point mutation) is a related name often used when dealing with structural problems (see [124, 280]).

While it is easily conceivable that a major deletion in a gene can lead to protein deficiency and consequently potentially life-threatening events, it is much more challenging to understand how a single amino acid substitution can cause disease, and this is even more intriguing, considering that proteins tend to be rather robust to site mutations [281, 282]. A key step in this process is of course to learn more about protein folding *in vivo* and quality control mechanisms. Important progresses have been made in understanding the functions of chaperones [283] and about protein folding (see for example [14, 15, 20]). The idea that inside a living organism, the mutant protein can be stabilized via contacts with surrounding molecules is also of importance (see also the concept of stabilization of proteins in confined spaces [284]). Another step is to understand the relationships between genetic and phenotypic variations and to assess the structural consequences of the mutations. Rules that are used to analyze the structural and functional effects of mutagenesis experiments can be applied here. Similarly, experimental structure analysis of protein variants and related biochemical data (e.g., calorimetry) can shed light on *in vivo* events.

In the following, we will consider only missense mutations (the most common type of genetic variations in humans) that can be analyzed at the protein structural level. Catalogues of naturally occurring mutations and associated molecular and clinical phenotypes have been reported for numerous proteins. Yet, it is not always clear if protein deficiency results from a given missense mutation, although absolute co-segregation between the mutation and the deficiency is certainly an important point. Mapping point

mutations onto 3D structures can not only clarify amino acid substitution-phenotype relationships but also provide significant insights about the roles of these residues on protein function, folding and stability [285-287]. Thus, combining clinical and structural studies should lead to a much broader understanding of some human diseases and help designing new treatments [2].

It has been shown, using site directed mutagenesis/conformation-specific monoclonal antibodies and X-ray crystallography, that residue substitutions (for instance in the core interior) by amino acids of slightly smaller/larger volumes (or even by charged residues) tend to destabilize the proteins but eventually do not alter significantly their 3D structures [129, 288-290]. Lack of important conformational changes illustrates the plasticity of protein cores, in which a new side chain can be tolerated via small local structural rearrangements. Yet, in some cases, structural shifts can propagate away from the site of mutation, sometime, up to 15-20 Å [195, 290, 291]. Cavity-filling mutations can stabilize a protein but alter its function if this latter is depending on slight movements within the core area that can be allowed by the presence of such internal cavities. These cavities can be lined by hydrophobic side chains (empty cavity) or by charged and/or polar residues (cavities filled with one or more water molecules) [292]. Mutations of residues lining these cavities can be detrimental to both stability and function. For instance, mutations that do not allow the presence of internal water molecules and H-bond networks can be destabilizing [293]. Creation of relatively large cavities within the protein core usually destabilizes a protein [293-295]. Large side chain deletions of buried and close packed residues within two structurally similar 3D structures can have little effect on stability of one molecule while being destabilizing for the other [296]. Possibly, the unaffected protein has greater intrinsic flexibility as compared to the destabilized molecule. Also, it has been proposed that residue depth relative to the protein surface correlates significantly better than solvent accessibility with regard to effects of mutations on protein stability and interactions [297].

Mutations of solvent exposed residues are often tolerated structurally but can cause significant functional disruption [298]. Substitution of solvent exposed residues by glycine can alter stability [293]. Proteins can undergo local folding/unfolding reactions and binding sites are often characterized by the presence of regions with low structural stability and regions with high structural stability [291, 299]. In enzymes, catalytic residues are usually (but not always) in regions with high structural stability (yet the important residues there can be found to be destabilizing). Such differences in stability can play a crucial role in the transmission of information from one region of a molecule toward (for instance) the catalytic site. Mutations in such regions could thus have dramatic effects on the function, while they would seem structurally tolerated. Salt-bridges can stabilize the native state and the unfolded states such that the net contribution to stability is small [300]. Charge-charge interactions can also contribute significantly to protein stability [301]. Salt-bridges/H-bonds can be destabilizing or only mildly stabilizing if isolated, but can be highly stabilizing when part of a larger network of

electrostatic interactions [172, 302-304]. Introduction of a charged residue in the protein core can induce behavior characteristic of molten globule states [290] or can be tolerated. Molten globule-like states can also occur when, for instance, a large residue (e.g., Phe) is replaced by a small amino acid (e.g., Ala) in the protein interior [305]. Regarding charged residues, the energetically unfavorable process of desolvation (e.g., charge burial into a hydrophobic cavity) will shift the pKa to favor the neutral form of the amino acids (upwards for acidic residues, downwards for basic residues). Coulombic interactions with neighboring charged or polar groups will generally favor the ionized form and can in some situations compensate for desolvation. A higher desolvation penalty is expected when charges are more localized (e.g., possibly Lys as compared to Glu). Also, Arg, Glu and Asp are able to form many hydrogen bonds and could be better tolerated in the protein core as compared to Lys. Polar residues inside the protein core can be stabilizing. Polar-to-hydrophobic substitutions at the surface of a protein can be tolerated in some regions and prohibited in other segments [306]. Unexpectedly, substituting solvent exposed charged residues by hydrophobic amino acids led to increased protein stability [307]. Mutations that change the configurational entropy of the unfolded state can modify protein stability [308, 309]. Mutations that shorten loops can stabilize a protein but be detrimental to its functions. Mutations of buried residues located on secondary structure elements can damage both the structure and the function. Mutations can change the folding pathway of proteins [131].

Considering the large range of effects possibly induced by (naturally occurring) point mutations, we define a set of simple rules aimed at facilitating structural analysis of protein variants (Table I). These rules follow the ones proposed by Wang and Moult [287] but new ones have been added. Naturally occurring point mutations can thus affect folding, stability, multimerization, ligand binding, catalysis, allosteric regulation, or post-translational modification.

Often, point mutations are only mildly destabilizing (around 2-5 kcal/mol), yet they can be highly destructive to function. Mutations that affect the stability of proteins seem to be the most common type as compared to destruction of binding sites or catalysis and others [287]. Unstable protein variants may be trapped inside the cells and rapidly cleared from the system.

As it is not possible to investigate experimentally all naturally occurring mutations, computer approaches can be used. Thus interactive structural analysis and molecular dynamics have been proven useful to analyze mutations prior to site directed mutagenesis, or to investigate the behavior of mutant proteins [116, 117, 120]. Along this line, investigation of wild-type and mutant Alzheimer's - amyloid peptides by combining NMR and molecular dynamics has been performed in order to understand better how a point mutation relates to disease [310]. Multiple sequence and structural alignments are also helpful in order to assess if residues are conserved, if the regions are flexible or rigid, or if spatially correlated mutations occur in a given area. Effective energy functions and inverse protein folding methods can give some insights [90, 92]. Approaches such as the MAGE/PROBE and PoPMuSiC methods can also be

Table I. Biostructural Pathology: Some Rules for Assigning the Effects of Missense Mutations on Molecular Functions, Folding and Stability

- ◆ Loss of hydrogen bond(s)
- ◆ Loss of burial of non-polar area on folding
- ◆ Over-exposure of non-polar residues (e.g., aggregation)
- ◆ Loss of salt-bridge(s) (assuming it is stabilizing, thus possibly part of a network and/or highly conserved between species, suggesting stabilization and/or specificity of the fold)
- ◆ Introduction of charged residues or polar residues in a hydrophobic pocket
- ◆ Introduction of large side chain in a tightly packed area
- ◆ Creation of an internal cavity or destruction of an internal cavity
- ◆ Substitution of a large solvent exposed residue by an amino acid with a smaller side chain that allows water molecules or ions to compete with internal H-bonds
- ◆ Disruption of metal binding sites
- ◆ Breakage of a disulfide bond (disulfide bonds can significantly stabilize proteins, possibly because of a decrease in the configurational chain entropy of the unfolded state)
- ◆ Apparition of a free Cys, which either disturbs normal disulfide bonding pattern or becomes exposed at the protein surface
- ◆ Backbone strain (e.g., Gly replaced by a large residue, any amino acid changed to Pro) either in loop structures or in secondary structure elements (destabilization of secondary structures)
- ◆ Destabilization of a protein multimer (e.g., mutation at the interface)
- ◆ Post-translational modification (e.g., disruption/creation of a N- or O-glycosylation or sulfation or phosphorylation site)
- ◆ Electrostatic repulsion via introduction of charged groups nearby like-charges
- ◆ Destruction of aromatic-aromatic interactions, of cation-pi electron interactions
- ◆ Destruction of a binding site or binding exosite
- ◆ Disruption of residues stabilizing the helix macrodipole
- ◆ Removal of residue involved in catalytic activity
- ◆ Amino acid changes away from the active site/binding site that alters the pKa of residues responsible of the catalytic activity or involved in intermolecular interaction
- ◆ Over-stabilization of a segment (e.g., extra Pro) or increased flexibility of a segment (e.g., extra Gly)
- ◆ Mutation of a residue located on an intra-molecular pathway that no longer allows propagation of a signal upon binding of ligand to distal regions (e.g., alteration of allosteric effects).

used for this purpose [311, 312]. Other structural analysis and theoretical studies have also been reported for evaluation of potential impacts of mutations on folding and/or stability (see for instance [195, 313, 314]).

3. ILLUSTRATIVE EXAMPLES: BLOOD COAGULATION PROTEINS

Cardiovascular diseases and associated blood coagulation disorders are leading causes of death worldwide. Blood coagulation involves more than 30 proteins that interact specifically with various degrees of affinity. Some of these molecules can bind transiently to membrane while others are membrane receptors. Series of catalytic reactions are taking place and the entire system is efficiently regulated under non-pathological situations [315]. However, hundreds of naturally occurring mutations (sometimes leading to complete deficiency) in proteins from the coagulation cascade and regulators have been reported, some being considered as risk factors for coagulation disorders, some being neutral. Investigation of the coagulation system via structural bioinformatics methods is therefore gratifying, not only for fundamental reasons, but also because understanding new

mechanisms helps in designing new diagnostic kits and more specific therapeutic compounds.

Numerous modeling studies have been reported, and we can only mention some of the most recent ones in the present review. Overall, study topics range from molecular dynamics of factors IXa/IX and Xa [316, 317], structure-based drug design for Xa and thrombin [318-320]; homology modeling of factor XIa and analysis of heparin binding [321], modeling of XIIa [322], modeling of the lectin-like domain of thrombomodulin [323] and suggested absence of calcium binding confirmed experimentally [324], docking of heparin-like molecules onto antithrombin [325], analysis of naturally occurring mutations in protein C or factor X or factor XIII or prothrombin [326-330], to modeling macromolecular complexes [331-333].

In the following paragraphs, we describe modeling studies related to C4b-binding protein (C4BP) and protein S (PS), protein C (PC) and the endothelial cell receptor (EPCR) for PC/activated protein C (APC), and factor V (FV) and factor VIII (FVIII). Such *in silico* investigations definitively help plan experiments, facilitate analysis of

naturally occurring mutations and give new insights about important molecular mechanisms.

3.1. Protein S – C4BP

Protein S (PS, 635 residues) is an anticoagulant non-enzymatic vitamin K-dependent glycoprotein that functions mainly as a cofactor to APC in the cleavage of coagulation FVa and FVIIIa. Protein S is structurally related to growth arrest specific factor 6 (Gas6), a ligand for some tyrosine kinase receptors [334]. A large amount of PS (about 60 %) is non-covalently associated and circulates with a molecule involved in the regulation of the complement system, C4b-binding protein (C4BP). Within this macromolecular complex, PS loses most of its anticoagulant properties. Protein S deficiency is a risk factor for thrombosis, and

numerous mutations in the PS gene have been reported [335]. Thus, understanding the roles of PS and its interactions with C4BP is of clinical importance, and inhibiting the PS-C4BP complex formation using small molecules could be beneficial to some patients suffering from thrombosis.

Protein S is a mosaic protein composed of multiple domains or modules (Fig. 1) [336]. Starting from the N-terminus, it contains a γ -carboxyglutamic acid-rich domain (or Gla domain important for interaction with the membrane, it binds calcium ions), a thrombin sensitive region (TSR), four epidermal growth factor (EGF)-like modules and a sex hormone binding globulin (SHBG)-like region. This last region has sequence homology (about 27 % sequence identity) with plasma sex hormone binding globulin protein (plasma SHBG), a molecule that binds steroid hormones

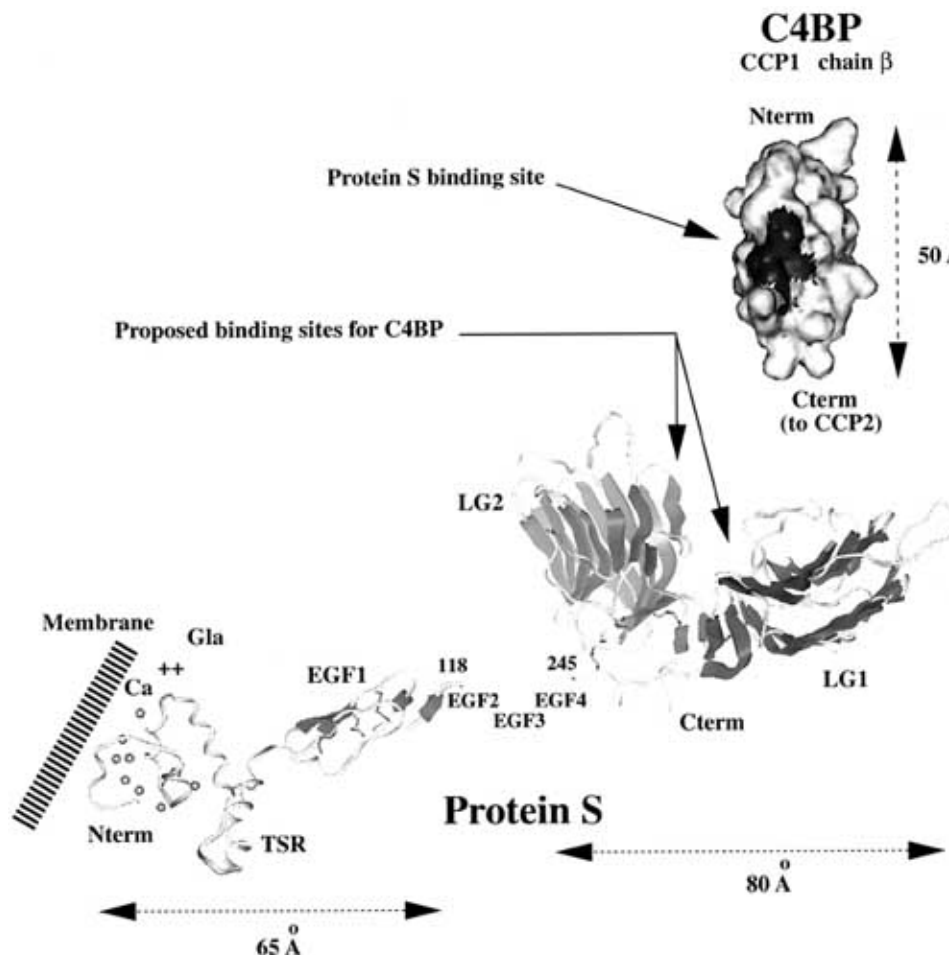


Fig. (1). Structural investigation of the C4BP-protein S interaction.

Models for regions of PS were developed in order to facilitate analysis of missense mutations, membrane binding and interaction with C4BP. At present, only the Gla-TSR-EGF1 region and the SHBG-like domain (two LG domains) have been built. The experimentally proposed binding sites for C4BP on the two LG domains have been investigated on the model structure, but additional mutagenesis studies are required to fully define the recognition surface. With the same scale as the one used for PS, the first CCP module of the C4BP γ -chain is shown (molecular surface). In dark, the cluster of hydrophobic residues proposed after structural analysis to be the key binding site for PS is shown. The importance of this patch has been fully validated via site directed mutagenesis.

[337]. The SHBG-like region of PS and plasma SHBG have been further subdivided using sensitive sequence comparison tools into two laminin G-type domains (LG domain, each of about 190 residues) [338, 339].

C4BP (about 4627 residues) is a spider-like multimeric glycoprotein composed of 6 or 7 α -chains, with or without a β -chain [337]. The α -chain and β -chain of C4BP are composed of 3 and 8 modules, respectively, of about 60 residues, called short consensus repeats (SCR) or CCP modules, further continued by a carboxyl-terminal region involved in the polymerization of the α - and β -chains [340, 341]. It has been suggested, after experimental studies, that the first CCP module of the α -chain contained a key interaction site for PS [342-344].

Models for the Gla-TSR-EGF1 region of PS have been developed [345, 346] to probe PS-APC and PS-cell membrane interactions [346, 347] and to analyze missense mutations [345, 348]. Models for this region of PS were built using sequence analysis, homology modeling, short molecular dynamics or Monte Carlo simulation with statistical potential and related X-ray and NMR template structures [349, 350]. Using site directed mutagenesis experiments and molecular modeling, a binding site for a calcium-dependent monoclonal antibody that does not inhibit interaction of PS with the cell membrane was localized around residues 49 and 52 of the TSR [346]. This information suggests that the Gla domain of PS has to be tilted with regard to the membrane plane in order to accommodate the large antibody molecule (Fig. 1). Also, despite the presence of several hydrophobic and/or aromatic residues within the TSR, it appeared that this segment of PS should contribute to the stabilization of the Gla domain but that the loop should not interact directly with the membrane. Some new experimental evidence supports these points [351]. Structural analysis of naturally occurring mutations mapping to this region of PS was carried out. One amino-acid substitution (Arg49His) within the TSR, was suggested to be associated with quantitative PS deficiency [352]. Investigation of the model and multiple sequence alignment showed that this residue is not conserved and apparently is solvent exposed. As such, structural data indicated that the substitution could be structurally tolerated. This hypothesis was confirmed via site directed mutagenesis experiments [348] and additional studies of co-segregation between plasma phenotypes and the mutation in the affected family [335]. This example illustrates how integration of modeling, biochemical and clinical studies help to understand the relationships between genetic mutations, molecular and clinical phenotypes.

Investigation of the SHBG-like region of PS has been hampered because of the lack of structural information about such domains. It was suggested that this region of PS could be subdivided into two LG domains [339]. However, because of the low sequence identity (about 10-15 %), it was difficult to define accurately the exact domain boundaries [353]. Multiple sequence alignments and secondary structure predictions were then carried out in order to gain some initial structural insights into the SHBG-like region of PS. Shortly after, the LG domains were merged to the pentraxin family [354]. The prediction of a 3D model using the pentraxin fold

for PS was, however, not possible at that time because of the extreme difficulty in generating accurate sequence to structure alignments. Clarifications came from X-ray studies of plasma SHBG, laminin LG domains and neurexin Ib [355-358]. Models for the SHBG-like region of PS were then developed using the Accelrys software package (San Diego, CA, USA) running on a Silicon Graphics R12000 O2 workstation (Mountain View, CA, USA) [359]. Secondary structure predictions over-estimated the α -helical content by about 10-15 % as compared to the structural model (mainly α -strands) while the overall location of the secondary structure elements was essentially correct [353, 359]. Such artifacts regarding over- or under-estimation of secondary structure elements is well known [65]. Proposed binding sites for C4BP on the surface of the SHBG-like region of PS were investigated on the model structure, but additional mutagenesis studies are required to fully characterize this interaction [359]. Naturally occurring point mutations were mapped onto the 3D model and analyzed [359, 360]. Overall, the mutations involved highly conserved and buried residues, localized within secondary structure elements or at the interface between the two LG domains. Some mutations destroyed disulfide bonds while others led to the burial of charged residues in the tightly packed hydrophobic core. Other mutations appeared to create a cavity. Some substitutions difficult to classify clinically could be assessed further in the 3D model. It is likely that many mutations destabilize (slightly) the structure rather than obstructing folding. Thus, subtle quality control systems should be able to detect even moderately unstable proteins and contribute to rapid clearance (e.g., degradation inside the cells) of these molecules.

Investigation of the C4BP-PS interaction is another example illustrating application of structural bioinformatics, potentially leading to development of new therapeutic compounds aiming at setting free more PS molecules. It was known that the C4BP α -chain was involved in PS binding [340]. A 3D model for this chain was developed and a solvent exposed hydrophobic patch was observed next to the junction with the second CCP [361]. Experimental characterization of the PS-C4BP interaction was then performed under increasing salt concentrations to confirm or reject the hypothesis resulting from structural analysis. It was indeed noticed that NaCl had very limited effects on the reaction, suggesting that electrostatic forces were not essential to the PS-C4BP interaction [362]. Theoretical screening [216] of the model molecular surface also highlighted the potential importance of this hydrophobic patch [362]. Further experimental evidence about the role of this hydrophobic surface came from mutagenesis studies, as substitutions of a few residues that were part of the hydrophobic cluster reduced dramatically the PS-C4BP interaction [363]. Such data are in agreement with the role of solvent exposed hydrophobic/aromatic patches discussed in the previous section for formation of a stable macromolecular complex. In contrast, a positively charged cluster was noticed in the C4BP α -chain model and was found to be essential for transient binding of complement protein C4b (the interaction was highly sensitive to salt concentration) [364, 365]. A binding surface displaying charged residues and able to form (weak) electrostatic contacts seems more appropriate than surface exposed hydrophobic clusters for

transient binding. These examples illustrate how CCP modules, via variation of their molecular surface, are able to bind very different proteins, with different affinities and high specificity. Small molecules could be designed rationally, based upon the modeling studies, to inhibit interactions between C4BP and PS or C4b.

3.2. Protein C – EPCR

A cascade of zymogen activation events is essential to the coagulation system and its regulation [366]. Protein C (PC; 419 residues), a multi-modular zymogen, is activated to an anticoagulant serine protease (activated PC or APC) by thrombin bound to thrombomodulin on the surface of endothelial cells. APC inactivates FVa and FVIIIa in the presence of its cofactor, PS [315]. From a structural standpoint, PC is a vitamin K-dependent plasma glycoprotein that has the same modular organization as many other blood coagulation enzymes (e.g., factor VII or factor IX). Protein C consists of a phospholipid binding -carboxyglutamic acid (Gla)-rich domain (which binds at least 7 calcium ions), two epidermal growth factor (EGF)-like domains and a serine protease domain (Fig. 2). The Gla-EGF1-EGF2 region constitutes the light chain while the serine protease (SP) domain is referred to as the heavy chain. These two chains are covalently linked via a disulfide bond that bridges the light chain to the heavy chain. An X-ray structure for Gla-domainless APC was reported in 1996 [367], but homology models for the SP domain (activated and zymogen forms) were developed prior to the experimental structure based upon related serine proteases [368, 369]. These models superimpose well onto the experimental structure and several important regions were already highlighted on these theoretical structures. Missense mutations causing protein C deficiency were investigated on these models [369, 370], and it was found that substitutions that led to complete protein deficiency essentially mapped inside the hydrophobic core. Models of the catalytic domain of early mammalian protein C were also reported [371]. More recently, models for full length APC and zymogen have been developed with the aim of understanding better the activation process by the complex thrombin-thrombomodulin [372-374]. The overall elongated shape of the PC/APC models was further cross-validated using size exclusion chromatography, which allowed evaluation of the Stokes radius, frictional ratio and axial ratio [373]. Potential binding sites at the surface of PC/APC using surface hydro-phobicity as a determinant of the preferred sites of intermolecular recognition were then reported, and interesting regions were noticed, for instance, on the first EGF domain.

The endothelial cell receptor (EPCR) for PC/APC is a 221 amino-acid residue long transmembrane glycoprotein. EPCR regulates the protein C anticoagulant pathway by binding PC/APC in a calcium-dependent manner and augmenting PC activation by the thrombin-thrombomodulin complex [375]. To facilitate future studies and to rationalize reported experimental data about this receptor, 3D models of human, bovine and mouse EPCR were developed using threading, comparative model building and secondary

structure prediction for the single transmembrane helix [376]. Clearly, the region preceding the transmembrane segment and the overall orientation of the receptor with regard to the membrane plane could not be predicted accurately. EPCR is related to CD1/MHC class I molecules (less than 30 % sequence identity) [375, 377]. It should consist of two domains, which are similar to the 1 and 2 domains of MHC class I molecules, whereas the 3 domain of MHC is replaced in EPCR by a transmembrane region followed by a short cytosolic tail. The 3 domain is important for the interaction with 2-microglobulin. Also, the 1 and 2 domains of CD1/MHC proteins are essential to the function of the molecules as they form a groove that usually binds short peptides. These two domains are composed of an eight-stranded antiparallel β -pleated sheet with two long antiparallel α -helices sitting on top. The distance between the helical segments dictates the width of the groove. In the EPCR models, the cleft appeared to be relatively narrow and was lined with hydrophobic/aromatic and polar residues, with a few charged amino acids. Analysis of the human EPCR model suggested, for instance, that the protein does not contain any calcium-binding pocket and that the four potential glycosylation sites are solvent exposed. As mentioned earlier, comparative analysis of related structures provides insights into the location and nature of potential binding sites. The structure and electrostatic properties of the three EPCR models were thus compared with the aim of identifying striking regions and recognition sites for PC/APC (Fig. 2). We assumed that the binding site for APC/PC at the surface of EPCR was conserved between species. Therefore, regions displaying important variations at the EPCR surface should not be involved in APC/PC binding. For instance, the fact that mouse EPCR presents a very acidic loop around position 106 suggests that this part of the receptor is not involved in APC/PC binding and thus, that site directed mutagenesis aiming at probing the interaction should be performed elsewhere. Keeping in mind that binding site locations can be conserved at the surface of structurally related proteins, it was suggested that a potential binding site for APC/PC could be present at the distal end of the two helical segments, as in the case of the neonatal Fc receptor [376]. As seen in (Fig. 2), this region is located opposite to the EPCR loop 106. The predicted binding site was confirmed via a combined modeling and site directed mutagenesis study [378].

Understanding the EPCR-APC/PC interaction is essential, as protein C is used as a therapeutic agent within hospital settings. Furthermore, engineering APC to a super-anticoagulant protein would most likely be beneficial to some pathology.

3.3. Factor V and Factor VIII

Two homologous non-enzymatic cofactors, factor V (FV) and factor VIII (FVIII) play important roles in the coagulation cascade, as readily illustrated by severe bleeding disorders occurring in patients deficient in FV and/or FVIII [379-381]. These two cofactors are essential to generation of thrombin molecules via formation of macromolecular

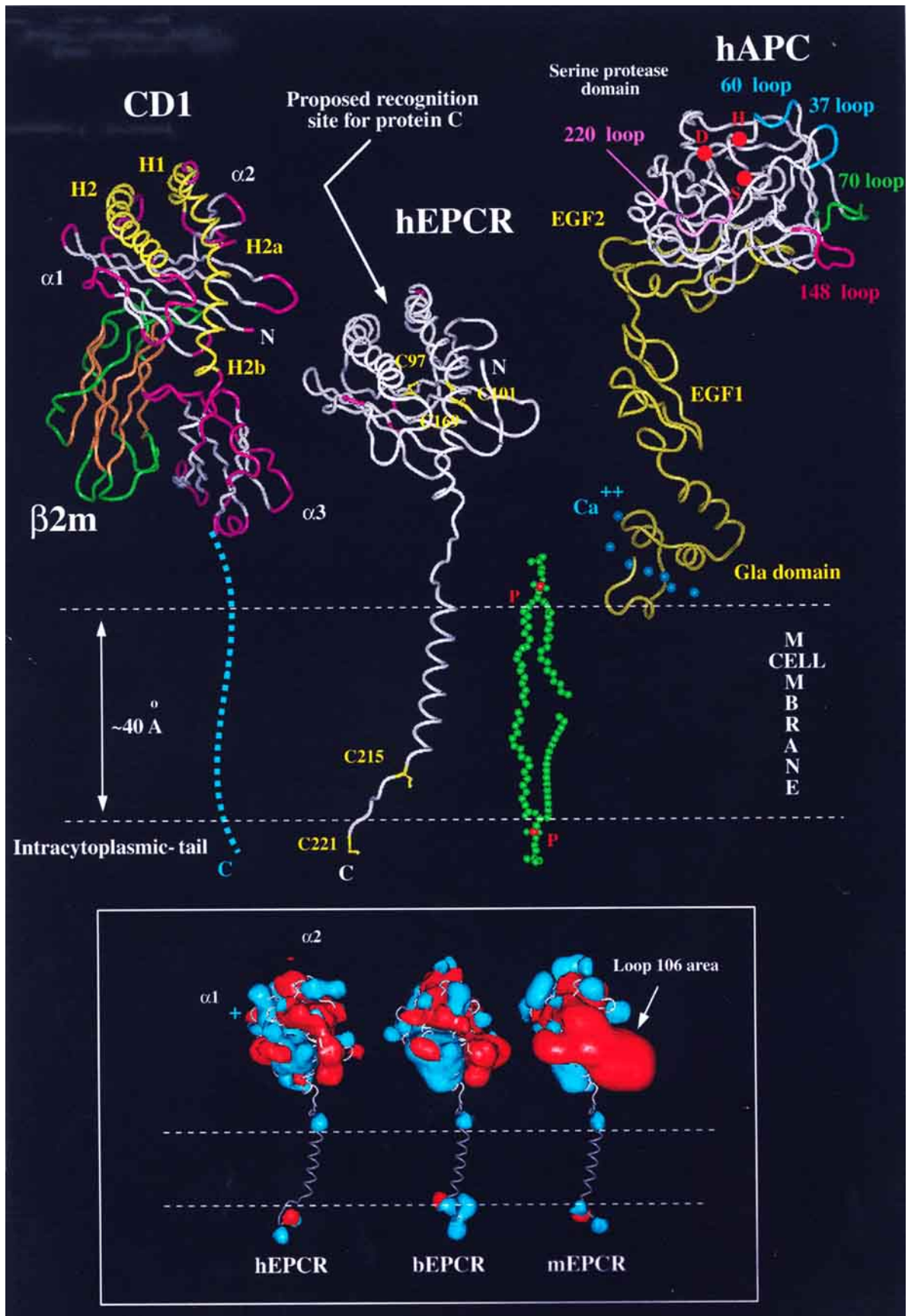


Fig. (2). Structural analysis of the protein C-EPCR interaction.

Legend (Fig. 2) contd....

The X-ray structures of mouse CD1 and 2M are shown (left). The CD1 helices are in yellow and the strands are white (orange in 2M). The structure of the CD1 transmembrane segment is not known, and is shown here as a dashed line (blue). Human EPCR model is presented in the middle with the Cys residues proposed to be involved in a disulfide bond yellow (C101-C169). The suggested free Cys (C97) is buried and is part of a β -strand. The N-term and C-term residues in CD1 and EPCR are noted N and C, respectively. Potentially glycosylated Asn residues in EPCR are shown in magenta. The model of human full length APC is presented (right) in order to show the relative size of the molecules. The serine protease (SP) domain is in white with some key surface loops being labeled and colored for orientation. The catalytic triad residues are in red (filled circles), and from left to right, they correspond to Asp 102, His 57 and Ser 195. The numbering for SP domain residues and loops follows the chymotrypsinogen nomenclature. The overall orientation of these molecules with respect to the membrane plane has not been experimentally proven yet, but is reasonable. A virtual membrane with a width of about 40 Å is displayed. Two phospholipid molecules were extracted from a lipid bilayer to aid in reading the figure. Phosphorus atoms are in red.

Fig. (2 inset). Electrostatic potential isosurfaces of the EPCR models.

The electrostatic iso-surfaces are displayed at a level of -1 (red) and +1 (blue) kcal/mol/e. In mouse EPCR, a region presenting a strikingly negative potential is observed. The transmembrane segment and part of the binding groove are neutral. The intracytoplasmic tail contains two to three Arg residues that could interact with negative groups present in the inner part of the phospholipid bilayer.

complexes (prothrombinase and tenase) [315], and understanding these reactions in details could help in designing new anticoagulant substances.

Starting from the N-terminus, FV and FVIII consist of domains A1, A2, B, A3, C1 and C2. The A domains have each about 300 residues, the B domain, heavily glycosylated, has about 900 residues, and each C domain has about 150 residues. The two molecules possess a heavy chain (domains A1, A2 and B) and a light chain (domains A3, C1 and C2). Activation by thrombin leads to the active cofactors, FVIIIa and FVa that lack the B domain. Binding to appropriate phospholipid membranes represents a key step allowing these two molecules to fully act as cofactors. FVIIIa and FVa are cleaved and inactivated by the complex APC-PS, thereby down-regulating further generation of thrombin and blocking the coagulation cascade.

The three A domains of FV and FVIII are homologous (about 40 % sequence identity) to the three A domains of ceruloplasmin (CP), the major transport protein for copper in plasma. Human CP is a single chain glycoprotein of 1046 amino acid residues. It belongs to a family of blue copper oxidases, with structural subunits based on the cupredoxin domain [382]. The cupredoxin fold is an eight-stranded Greek key β -barrel that was observed first in plastocyanin and azurin.

Theoretical models for the A domains of FVIII [383, 384] and FV [385, 386] were recently reported using the well established comparative model building method based on the X-ray structure of CP [387]. Overall, the dimensions of this region of the FVIII and FV molecules, as evaluated from computer modeling, are consistent with recent electron microscopy studies [388]. Numerous experimental data could be mapped onto the 3D structure of the A domains, including two cleavage sites for APC. Some mutations in the FV gene associated with FV deficiency were evaluated in the 3D structure [389, 390]. Similarly, mutations mapping in the A domains of FVIII have been investigated [383, 391-394]. The A domain model has also been used recently to

propose an initial structure of the factor IXa:factor VIIIa complex [395]. Possibly, analysis of the 3D structure of human FVIII and related species could help reduce the antigenicity of factor VIII toward inhibitory antibodies.

The FV and FVIII C domains belong to the discoidin domain family [396], for which no structural information was known until recent crystallographic studies [397, 398]. However, sensitive sequence comparisons merged the discoidin domain with the D1 domain of galactose oxidase (GOase) [399] in spite of sequence identities ranging from about 8 to 12 % [7]. In parallel modeling studies using the threading method reported by Fischer *et al.* [400] and a comparative model building protocol [42] were performed, and found to be in agreement with the sequence analysis investigation of Baumgartner *et al.* [7, 401]. In order to evaluate the overall accuracy of the predicted structure, several cysteine residues known to form disulfide bonds in several discoidin domains were mapped onto the model and found sufficiently close in space to make such bonds. Similarly, glycosylated residues in several of these domains were seen at the surface of the model, thus confirming the structural predictions. Because it was known that the C2 domain was important for membrane binding, screening the surface of the theoretical model highlighted the presence of a solvent exposed hydrophobic loop (loop 2060 in factor V) and clusters of positively charged residues [401]. This loop was thus suggested to be of importance for membrane binding, for both FV and FVIII. Another modeling study suggested the C domains of FV and FVIII to adopt the fold of the D1 domain of GOase, and underlined the importance of some loops for membrane binding [402, 403].

However, these modeling studies were in disagreement with NMR data indicating that a segment of the discoidin domain of FVIII adopted a helical structure in the presence of lipids [404]. This region was in fact found to be a loop and beta strand in the X-ray structure of FVIII, and was unlikely to become an α -helix upon membrane binding [397]. Comparison of the theoretical model and X-ray structure for the C2 domain of FV shows that the model structure is

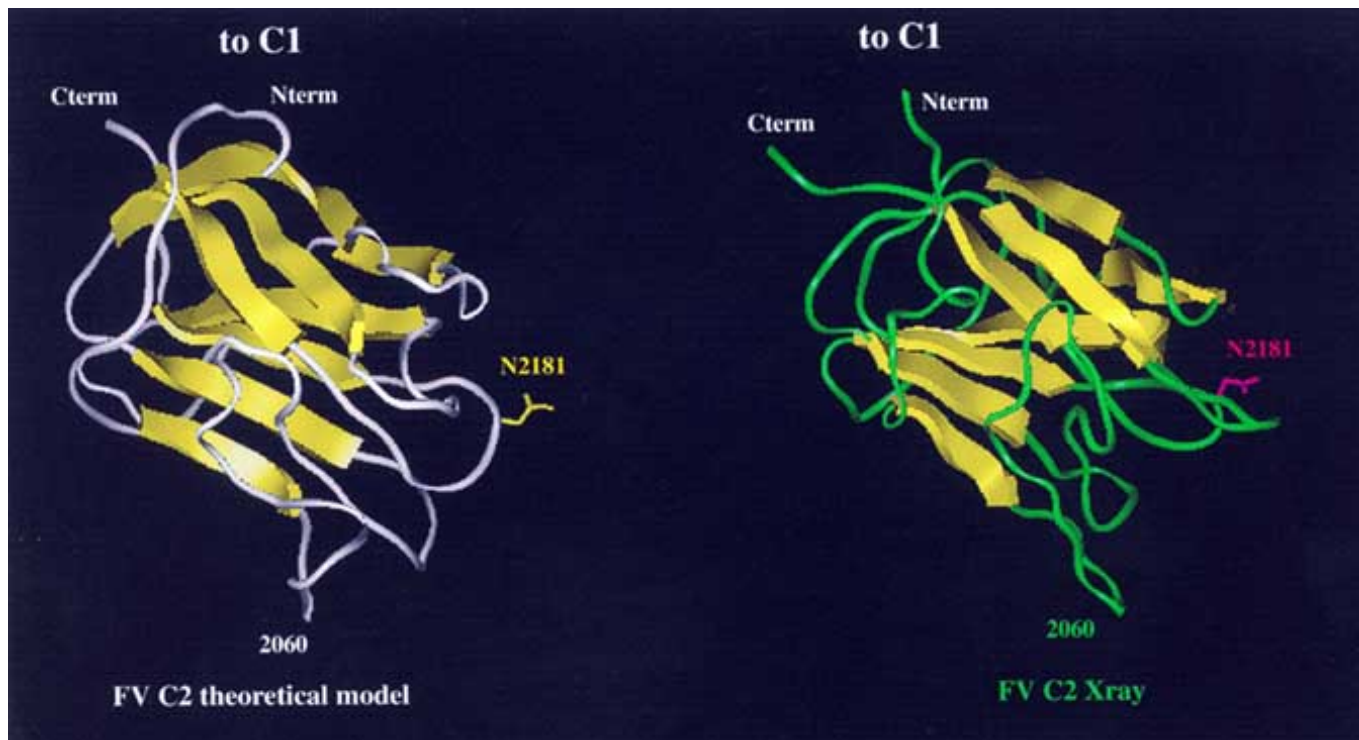


Fig. (3). Theoretical model and X-ray structure for the C2 domain of factor V.

The theoretical model is shown to the left, while the X-ray structure of the domain, reported one year after the predicted structure, is presented in the right part of the panel. Glycosylated Asn 2181 is displayed to facilitate orientation. As expected, some differences are noted in loop regions when comparing the model and the X-ray structure. However, the theoretical model was sufficiently accurate to define key functional regions such as the 2060-loop. This segment, containing two tryptophan residues, was shown experimentally, at a later stage, to be important for membrane binding

essentially correct (Fig. 3). The factor V C2 model was used to design and analyze site directed mutagenesis experiments aimed at investigating the role of glycosylation at residue Asn2181, and possible functions of the loop 2060 with regard to membrane binding [405, 406]. Models for the C1 domains of FV and FVIII based on the above mentioned X-ray structures were recently developed and used to investigate naturally occurring point mutations and membrane binding processes [407-410]. A model for membrane-bound FVIII using a 3D molecular envelope derived from electron microscopy and molecular modeling has been reported, showing for the first time the overall organization of the A and C domains with regard to the membrane plane [411]. This structure provides direct evidence for the insertions of some loops of the C2 domain into the lipids. The other domains were not found to contact the lipid surface directly, but flexibility between the C domains may allow such movements (e.g., upon protein-protein interactions as within the intrinsic tenase complex).

These recently reported structural information open new avenues to study both FV and FVIII, and there can be little doubt that our understanding of the "tenase" and "prothrombinase" complexes will increase significantly over the course of the next decade.

CONCLUSION

There is a growing trend towards the use of computational procedures to model laboratory results before actual physical experimentation. Indeed, protein modeling methods and *in silico* investigations yield information complementary to NMR spectroscopy, X-ray crystallography, biochemistry and molecular biology. These computational methods can be used to predict 3D structures, potential protein-protein interaction sites and other molecular functions. Progress in the development of more accurate energy functions should facilitate structural predictions and investigations of the dynamics and alternative states of proteins. As docking procedures become more accurate, it should be easier to develop models for macromolecular complexes. Structure-based ligand design protocols have been successful in numerous cases and should facilitate generation of novel pharmacologically active molecules. Computer modeling tools also allow for the analysis of both theoretical and experimental 3D structures, either alone or in combination with mutagenesis and other biophysical methods such as fluorescence resonance energy transfer, small angle X-ray scattering and electron microscopy. With regard to cardiovascular diseases, structural bioinformatics concepts/methods help to plan and rationalize experiments and simplify understanding of some molecular mechanisms.

Because cardiovascular diseases are leading causes of death worldwide and are among the first causes of severe acquired disability, it is important to associate experimental and theoretical approaches to generate new information about some key molecular mechanisms and ultimately to develop new chemotherapeutically active reagents. Better understanding of conformational diseases and how point mutations relate to folding/stability problems and functions represents another important step that should contribute to the development of new therapeutic compounds.

ACKNOWLEDGEMENTS

This work was supported by a grant from La Fondation pour La Recherche Médicale. We are grateful to Dr. Kathleen P. Pratt for critical reading of the manuscript. We thank Dr. Martine Aiach for reading the manuscript at an early stage of writing. The electronic journal MMCC (Molecular Modeling & Computational Chemistry)-Results by Dr. D. Busath and collaborators has been very helpful to identify research papers and reviews focusing on bioinformatics. We thank Dr. Boris P. Atanasov for interesting discussions about electrostatics and Drs. Johan Stenflo and Björn Dahlbäck for helpful comments about blood coagulation proteins over the years.

REFERENCES

- [1] Thornton, J. M. (2001) *Protein Sci.* 10, 3-11.
- [2] Perutz, M. *Protein structure, new approaches to disease and therapy*. New York, W. H. Freeman and Company, 1992.
- [3] Anthonsen, H. W.; Baptista, A.; Drablos, F.; Martel, P. and Petersen, S. B. (1994) *J. Biotechnol.* 36, 185-220.
- [4] Michielin, O.; Luescher, I. and Karplus, M. (2000) *J. Mol. Biol.* 300, 1205-35.
- [5] Gordon, D. B.; Marshall, S. A. and Mayo, S. L. (1999) *Curr. Opin. Struct. Biol.* 9, 509-13.
- [6] Sowa, M. E.; He, W.; Slep, K. C.; Kercher, M. A.; Lichtarge, O. and Wensel, T. G. (2001) *Nat. Struct. Biol.* 8, 234-7.
- [7] Baumgartner, S.; Hofmann, K.; Chiquet-Ehrismann, R. and Bucher, P. (1998) *Protein Sci.* 7, 1626-31.
- [8] Landgraft, R.; Xenarios, L. and Eisenberg, D. (2001) *J. Mol. Biol.* 307, 1487-502.
- [9] Eisenberg, D.; Marcotte, E. M.; Xenarios, I. and Yeates, T. O. (2000) *Nature* 405, 823-6.
- [10] Marcotte, E. M.; Pellegrini, M.; Ng, H. L.; Rice, D. W.; Yeates, T. O. and Eisenberg, D. (1999) *Science* 285, 751-3.
- [11] Pellegrini, M.; Marcotte, E. M.; Thompson, M. J.; Eisenberg, D. and Yeates, T. O. (1999) *Proc. Natl. Acad. Sci. USA* 96, 4285-8.
- [12] Jager, M.; Nguyen, H.; Crane, J. C.; Kelly, J. W. and Gruebele, M. (2001) *J. Mol. Biol.* 311, 373-93.
- [13] Dill, K. A. (1990) *Biochemistry* 29, 7133-55.
- [14] Dill, K. A. and Chan, H. S. (1997) *Nat. Struct. Biol.* 4, 10-9.
- [15] Nolting, B. and Andert, K. (2000) *Proteins* 41, 288-98.
- [16] Lazaridis, T. and Karplus, M. (1997) *Science* 278, 1928-31.
- [17] Moulton, J. and Melamud, E. (2000) *Curr. Opin. Struct. Biol.* 10, 384-9.
- [18] Kuhlman, B. and Baker, D. (2000) *Proc. Natl. Acad. Sci. USA* 97, 10383-8.
- [19] Wallqvist, A.; Smythers, G. W. and Covell, D. G. (1997) *Protein Sci.* 6, 1627-42.
- [20] Honig, B. (1999) *J. Mol. Biol.* 293, 283-93.
- [21] Baker, D. and Agard, D. A. (1994) *Biochemistry* 33, 7505-9.
- [22] Thirumalai, D.; Klimov, D. K. and Woodson, S. A. (1997) *Theor. Chem. Acc.* 96, 14-22.
- [23] Dokholyan, N. V. and Shakhnovich, E. I. (2001) *J. Mol. Biol.* 312, 289-307.
- [24] Murphy, K. P.; Bhakuni, V.; Xie, D. and Freire, E. (1992) *J. Mol. Biol.* 227, 293-306.
- [25] Sheinerman, F. B. and Brooks, C. L. (1998) *Proc. Natl. Acad. Sci. USA* 95, 1562-7.
- [26] Vorobjev, Y. N. and Hermans, J. (2001) *Protein. Sci.* 10, 2498-506.
- [27] Vorobjev, Y. N.; Almagro, J. C. and Hermans, J. (1998) *Proteins* 32, 399-413.
- [28] Friedberg, I. and Margalit, H. (2002) *Protein Sci.* 11, 350-60.
- [29] Holm, L. and Sander, C. (1995) *Trends Biochem. Sci.* 20, 478-80.
- [30] Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N. and Bourne, P. E. (2000) *Nucleic Acids. Res.* 28, 235-42.
- [31] Marcotte, E. D. (2000) *Curr. Opin. Struct. Biol.* 10, 359-65.
- [32] Teichmann, S. A.; Murzin, A. G. and Chothia, C. (2001) *Curr. Opin. Struct. Biol.* 11, 354-63.
- [33] Xenarios, I.; Salwinski, L.; Duan, X. J.; Higney, P.; Kim, S. M. and Eisenberg, D. (2002) *Nucleic Acids Res.* 30, 303-5.
- [34] Eisenberg, D. (1999) *Nature* 402, 83-6.
- [35] Walhout, A. J.; Sordella, R.; Lu, X.; Hartley, J. L.; Temple, G. F.; Brasch, M. A.; Thierry-Mieg, N. and Vidal, M. (2000) *Science* 287, 116-22.
- [36] Wojcik, J. and Schachter, V. (2001) *Bioinformatics* 1, 1-10.
- [37] Spengler, S. J. (2000) *Science* 287, 1221-3.
- [38] Kumar, S. and Nussinov, R. (2001) *Cell. Mol. Life. Sci.* 58, 1-18.

- [39] Kikkawa, M.; Sablin, E. P.; Okada, Y.; Yajima, H.; Fletterick, R. J. and Hirokawa, N. (2001) *Nature* 411, 439-45.
- [40] Osguthorpe, D. J. (2000) *Curr. Opin. Struct. Biol.* 10, 146-52.
- [41] Samudrala, R.; Xia, Y.; Huang, E. and Levitt, M. (1999) *Proteins Suppl* 3, 194-8.
- [42] Greer, J. (1991) *Methods Enzymol.* 202, 239-52.
- [43] Bajorath, J.; Stenkamp, R. and Aruffo, A. (1993) *Protein Sci.* 2, 1798-810.
- [44] Blundell, T. L.; Sibanda, B. L.; Sternberg, M. J. E. and Thornton, J. M. (1987) *Nature* 326, 347-52.
- [45] Maggio, E. T. and Ramnarayan, K. (2001) *Trends Biotechnol.* 19, 266-72.
- [46] Hao, M. and Scheraga, H. A. (1999) *Curr. Opin. Struct. Biol.* 9, 184-8.
- [47] Baker, D. and Sali, A. (2001) *Science* 294, 93-6.
- [48] Simons, K. T.; Bonneau, R.; Ruczinski, I. and Baker, D. (1999) *Proteins Suppl* 3, 171-6.
- [49] Mirny, L. A.; Abkevich, V. I. and Shakhnovich, E. I. (1998) *Proc. Natl. Acad. Sci. USA* 95, 4976-81.
- [50] Skolnick, J.; Kolinski, A. and Ortiz, A. (1998) *J. Biomol. Struct. Dyn.* 16, 381-96.
- [51] Klimov, D. K. and Thirumalai, D. (1998) *Fold. Des.* 3, 127-39.
- [52] Pedersen, J. T. and Moulton, J. (1996) *Curr. Opin. Struct. Biol.* 6, 227-31.
- [53] Abagyan, R. and Totrov, M. (1994) *J. Mol. Biol.* 235, 983-1002.
- [54] Daura, X.; Jaun, B.; Seebach, D.; vanGunsteren, W. F. and Mark, A. E. (1998) *J. Mol. Biol.* 280, 925-32.
- [55] Duan, Y. and Kollman, P. A. (1998) *Science* 282, 740-4.
- [56] Gillis, D. and Rooman, M. (2001) *Proteins* 42, 164-76.
- [57] Ishikawa, K.; Yue, K. and Dill, K. A. (1999) *Protein Sci.* 8, 716-21.
- [58] Lee, M. R.; Baker, D. and Kollman, P. A. (2001) *J. Am. Chem. Soc.* 123, 1040-6.
- [59] Lee, M. R.; Tsai, J.; Baker, D. and Kollman, P. A. (2001) *J. Mol. Biol.* 313, 417-30.
- [60] Lee, M. R. and Kollman, P. A. (2001) *Structure* 9, 905-16.
- [61] Fariselli, P.; Riccobelli, P. and Casadio, R. (1999) *Proteins* 36, 340-6.
- [62] Mucchielli-Giorgi, M. H.; Hazout, S. and Tuffery, P. (2002) *Proteins* 46, 243-9.
- [63] Chothia, C. and Lesk, A. M. (1986) *EMBO J.* 5, 823-6.
- [64] Martin, A. C. R.; MacArthur, M. W. and Thornton, J. M. (1997) *Proteins Suppl* 1, 14-28.
- [65] Branden, C. I. and Tooze, J. *Introduction to Protein Structure* New York, Garland Publishing, 1999.
- [66] Amodeo, P.; Fraternali, F.; Lesk, A. M. and Pastore, A. (2001) *J. Mol. Biol.* 311, 283-96.
- [67] Sali, A. (1998) *Nat. Struct. Biol.* 5, 1029-32.
- [68] Bajorath, J. (1998) *J. Biol. Chem.* 273, 24603-9.
- [69] Sanchez, R. and Sali, A. (1997) *Curr. Opin. Struct. Biol.* 7, 206-14.
- [70] Yang, A. and Honig, B. (1999) *Proteins Suppl* 3, 66-72.
- [71] Xu, D.; Xu, Y. and Uberbacher, E. D. (2000) *Current Protein and Peptide Science* 1, 1-21.
- [72] Notredame, C.; Higgins, D. G. and Heringa, J. (2000) *J. Mol. Biol.* 302, 205-17.
- [73] Russel, R. B.; Saqi, M. A. S.; Bates, P. A.; Sayle, R. A. and Stenberg, M. J. E. (1998) *Prot. Eng.* 11, 1-9.
- [74] deBakker, P. I. W.; Bateman, A.; Burke, D. F.; Miguel, R. N.; Mizuguchi, K.; Shi, J. and Blundell, T. L. (2001) *Bioinformatics* 17, 748-9.
- [75] Shi, J.; Blundell, T. L. and Mizuguchi, K. (2001) *J. Mol. Biol.* 310, 243-57.
- [76] Miguel, R. N.; Shi, J. and Mizuguchi, K. Protein fold recognition and comparative modelling using HOMSTRAD, JOY and FUGUE. *Protein Structure prediction: bioinformatic approach* La Jolla, International University Line, 2002, in press.
- [77] Fischer, D.; Barret, C.; Bryson, K.; Elofsson, A.; Godzik, A.; Jones, D.; Karplus, K. J.; Kelley, L. A.; MacCallum, R. M.; Pawowski, K.; Rost, B.; Rychlewski, L. and Sternberg, M. (1999) *Proteins Suppl* 3, 209-17.
- [78] Chung, S. Y. and Subbiah, S. (1996) *Structure* 4, 1123-7.
- [79] Fischer, D.; Rice, D.; Bowie, J. U. and Eisenberg, D. (1996) *The FASEB J.* 10, 126-36.
- [80] Sippl, M. J. and Weitckus, S. (1992) *Proteins* 13, 258-71.
- [81] Jones, D. T. and Thornton, J. M. (1996) *Curr. Opin. Struct. Biol.* 6, 210-6.
- [82] Gibrat, J.-F.; Madej, T. and Bryant, S. H. (1996) *Curr. Opin. Struct. Biol.* 6, 377-85.
- [83] Rice, D. W. and Eisenberg, D. (1997) *J. Mol. Biol.* 267, 1026-38.
- [84] Duan, J.; Dahlback, B. and Villoutreix, B. O. (2001) *FEBS Letters* 499, 127-32.
- [85] Kolinski, A.; Betancourt, M. R.; Kihara, D.; Rotkiewicz, P. and Skolnick, J. (2001) *Proteins* 44, 133-49.
- [86] Rapp, C. S. and Friesner, R. A. (1999) *Proteins* 35, 173-83.
- [87] Burke, D. F.; Deane, C. M. and Blundell, T. L. (2000) *Bioinformatics* 16, 513-9.
- [88] Deane, C. M. and Blundell, T. L. (2001) *Protein Sci.* 10, 599-612.

- [89] Fiser, A.; Do, R. K. G. and Sali, A. (2000) *Protein Sci.* 9, 1753-73.
- [90] Lazaridis, T. and Karplus, M. (2000) *Curr. Opin. Struct. Biol.* 10, 139-45.
- [91] Crippen, G. M. (2001) *J. Mol. Biol.* 306, 565-73.
- [92] Luthy, R.; Bowie, J. U. and Eisenberg, D. (1992) *Nature* 356, 83-5.
- [93] Sippl, M. J. (1995) *Curr. Opin. Struct. Biol.* 5, 229-35.
- [94] Melo, F.; Sanchez, R. and Sali, A. (2002) *Protein Sci.* 11, 430-48.
- [95] Liang, J. and Dill, K. A. (2001) *Biophys. J.* 81, 751-66.
- [96] Karplus, M. and Petsko, G. A. (1990) *Nature* 347, 631-9.
- [97] Schlick, T. (2001) *Structure* 9, R45-R53.
- [98] Field, M. J. A. *practical introduction to the simulation of molecular systems*, Cambridge University Press, UK, 1999.
- [99] Daggett, V. (2000) *Curr. Opin. Struct. Biol.* 10, 160-4.
- [100] Caves, L. S.; Evanseck, J. D. and Karplus, M. (1998) *Protein Sci.* 7, 649-66.
- [101] Hunenberger, P. H. and McCammon, J. A. (1999) *Biophys. Chem.* 78, 69-88.
- [102] Schutz, C. N. and Warshel, A. (2001) *Proteins* 44, 400-17.
- [103] Darden, T.; Perera, L.; Li, L. and Pedersen, L. (1999) *Structure* 7, R55-R60.
- [104] Krueger, B. P. and Kollman, P. A. (2001) *Proteins* 44, 4-15.
- [105] David, L.; Luo, R. and Gilson, M. K. (2000) *J. Comp. Chem.* 21, 295-309.
- [106] Bashford, D. and Case, D. A. (2000) *Annu. Rev. Phys. Chem.* 51, 129-52.
- [107] Dominy, B. N. and Brooks, C. L. (1999) *J. Phys. Chem. B* 103, 3765-73.
- [108] Still, W. C.; Tempczyk, A.; Hawley, R. C. and Hendrickson, T. (1990) *J. Am. Chem. Soc.* 112, 6127-9.
- [109] Calimet, N.; Schaefer, M. and Simonson, T. (2001) *Proteins* 45, 144-58.
- [110] Chun, H. M.; Padilla, C. E.; Chin, D. N.; Watanabe, M.; Karlov, V. I.; Alper, H. E.; Soosaar, K.; Blair, K. B.; Becker, O. M.; Caves, L. S.; Nagle, R.; Haney, D. N. and Farmer, B. L. (2000) *J. Comp. Chem.* 21, 159-84.
- [111] Sommer, I. and Brimacombe, R. (2001) *J. Comp. Chem.* 22, 407-17.
- [112] Bahar, I.; Erman, B.; Jernigan, R. L.; Atilgan, A. R. and Covell, D. G. (1999) *J. Mol. Biol.* 285, 1023-37.
- [113] Berendsen, H. J. C. and Hayward, S. (2000) *Curr. Opin. Struct. Biol.* 10, 165-9.
- [114] Madrid, M.; Jacobo-Molina, A.; Ding, J. and Arnold, E. (1999) *Proteins* 35, 332-7.
- [115] Amadei, A.; Linssen, A. B. M. and Berendsen, H. J. C. (1993) *Proteins* 17, 412-25.
- [116] Voordijk, S.; Hansson, T.; Hilvert, D. and vanGunsteren, W. F. (2000) *J. Mol. Biol.* 300, 963-73.
- [117] Cregut, D. and Serrano, L. (1999) *Protein Sci.* 8, 271-82.
- [118] Lee, J.; Lee, K. and Shin, S. (2000) *Biophys. J.* 78, 1665-71.
- [119] Elmore, D. E. and Dougherty, D. A. (2001) *Biophys. J.* 81, 1345-59.
- [120] Roche, O. and Field, M. J. (2001) *Proteins* 45, 237-40.
- [121] Mayor, U.; Johnson, C. M.; Daggett, V. and Fersht, A. R. (2000) *Proc. Natl. Acad. Sci. USA* 97, 13518-22.
- [122] Sham, Y. Y.; Ma, B.; Tsai, C.-J. and Nussinov, R. (2002) *Proteins* 46, 308-20.
- [123] Alonso, D. O. V.; DeArmond, S. J.; Cohen, F. E. and Daggett, V. (2001) *Proc. Natl. Acad. Sci. USA* 98, 2985-9.
- [124] Ferreira, S. T. and Felice, F. G. d. (2001) *FEBS Lett.* 24916, 1-6.
- [125] Massi, F.; Peng, J. W.; Lee, J. P. and Straub, J. E. (2001) *Biophysical J.* 80, 31-44.
- [126] Jacobs, D. J.; Rader, A. J. and Kuhn, L. A. (2001) *Proteins* 44, 150-65.
- [127] Elcock, A. H. (2001) *J. Mol. Biol.* 312, 885-96.
- [128] Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J. and McCammon, J. A. (2001) *Proc. Natl. Acad. Sci. USA* 98, 10037-41.
- [129] Sanchez-Ruiz, J. M. and Makhatadze, G. I. (2001) *TRENDS Biotech.* 19, 132-5.
- [130] Whitten, S. T. and Garcia-Moreno, B. E. (2000) *Biochemistry* 39, 14292-304.
- [131] Garcia, C.; Nishimura, C.; Cavagnero, S.; Dyson, H. J. and Wright, P. E. (2000) *Biochemistry* 39, 11227-37.
- [132] Strop, P. and Mayo, S. L. (2000) *Biochemistry* 39, 1251-5.
- [133] Warshel, A. (1998) *J. Biol. Chem.* 273, 27035-8.
- [134] Villa, J.; Strajbl, M.; Glennon, T. M.; Sham, Y. Y.; Chu, Z. T. and Warshel, A. (2000) *Proc. Natl. Acad. Sci. USA* 97, 11899-904.
- [135] Honig, B. and Nicholls, A. (1995) *Science* 268, 1144-9.
- [136] Warshel, A. and Åqvist, J. (1991) *Annu. Rev. Biophys. Chem.* 20, 267-98.
- [137] Bashford, D. (1991) *Curr. Opin. Struct. Biol.* 1, 175-84.
- [138] Gilson, M. K. and Honig, B. (1988) *Proteins* 4, 7-18.
- [139] Wade, R. C.; Gabdouline, R. R.; Lüdemann, S. K. and Lounnas, V. (1998) *Proc. Natl. Acad. Sci. USA* 95, 5942-9.
- [140] Laberge, M. (1998) *Biochim. Biophys. Acta* 1386, 305-30.
- [141] daSilva, F. L. B.; Jonsson, B. and Penfold, R. (2001) *Protein Sci.* 10, 1415-25.

- [142] Zhang, L. Y.; Gallicchio, E.; Friesner, R. A. and Levy, R. M. (2001) *J. Comp. Chem.* 22, 591-607.
- [143] Simonson, T. (2001) *Curr. Opin. Struct. Biol.* 11, 243-52.
- [144] Nielsen, J. E.; Andersen, K. V.; Honig, B.; Hoof, R. W. W.; Klebe, G.; Vriend, G. and Wade, R. C. (1999) *Protein Eng.* 12, 657-62.
- [145] Havranek, J. J. and Harbury, P. B. (1999) *Proc. Natl. Acad. Sci. USA* 96, 11145-50.
- [146] Elcock, A. H. (1999) *J. Mol. Biol.* 294, 1051-62.
- [147] Bashford, D. and Karplus, M. (1990) *Biochemistry* 29, 10219-25.
- [148] Warwicker, J. (1999) *Protein Sci.* 8, 418-25.
- [149] Lounnas, V. and Wade, R. C. (1997) *Biochemistry* 36, 5402-17.
- [150] Sandberg, L. and Edholm, O. (1999) *Proteins* 36, 474-83.
- [151] Karshikoff, A. (1995) *Prot. Eng.* 8, 243-8.
- [152] Antosiewicz, J.; Briggs, J. M.; Elcock, A. H.; Gilson, M. K. and McCammon, J. A. (1996) *J. Comp. Chem.* 17, 1633-44.
- [153] Mehler, E. L. and Guarnieri, F. (1999) *Biophys. J.* 75, 3-22.
- [154] Miteva, M.; Andersson, M.; Karshikoff, A. and Otting, G. (1999) *FEBS Lett.* 462, 155-8.
- [155] Alexov, E. G. and Gunner, M. R. (1997) *Biophys. J.* 74, 2075-93.
- [156] Koumanov, A.; Karshikoff, A.; Friis, E. P. and Borchert, T. V. (2001) *J. Phys. Chem. B.* 105, 9339-44.
- [157] Spassov, V. Z.; Luecke, H.; Gerwert, K. and Bashford, D. (2001) *J. Mol. Biol.* 312, 203-19.
- [158] Koumanov, A.; Ruterjans, H. and Karshikoff, A. (2002) *Proteins* 46, 85-96.
- [159] Kao, Y.-H.; Fitch, C. A.; Bhattacharya, S.; Sarkisian, C. J.; Lecomte, J. T. J. and Garcia-Moreno, B. E. (2000) *Biophys. J.* 79, 1637-54.
- [160] Simonson, T. and Perahia, D. (1995) *Comp. Phys. Com.* 91, 291-303.
- [161] Pitera, J. W.; Falta, M. and vanGunsteren, W. F. (2001) *Biophys. J.* 80, 2546-56.
- [162] Dwyer, J. J.; Gittis, A. G.; Karp, D. A.; Lattman, E. E.; Spencer, D. S.; Stites, W. E. and Garcia-Moreno, B. E. (2000) *Biophys. J.* 79, 1610-20.
- [163] Rocchia, W.; Alexov, E. and Honig, B. (2001) *J. Phys. Chem. B.* 105, 6507-14.
- [164] Tanford, C. and Kirkwood, J. G. (1957) *J. Am. Chem. Soc.* 79, 5333-9.
- [165] Spassov, V. Z.; Karshikov, A. D. and Atanasov, B. P. (1989) *Biochim. Biophys. Acta* 999, 1-6.
- [166] Villoutreix, B. O.; Spassov, V. Z.; Atanasov, B. P.; Herve, G. and Ladjimi, M. M. (1994) *Proteins* 19, 230-43.
- [167] Neves-Petersen, M. T.; Petersen, E. I.; Fojan, P.; Noronha, M.; Madsen, R. G. and Petersen, S. B. (2001) *J. Biotech.* 87, 225-54.
- [168] Ibarra-Molero, B.; Loladze, V. V.; Makhatadze, G. I. and Sanchez-Ruiz, J. M. (1999) *Biochemistry* 38, 8138-49.
- [169] Dennis, S.; Camacho, C. J. and Vadja, S. (2000) *Proteins* 38, 176-88.
- [170] Lee, L.-P. and Tidor, B. (2001) *Protein Sci.* 10, 362-77.
- [171] Ullmann, G. M.; Hauswald, M.; Jensen, A. and Knapp, E.-W. (2000) *Proteins* 38, 301-9.
- [172] Hendsch, Z. S. and Tidor, B. (1994) *Protein Sci.* 3, 211-26.
- [173] Wimley, W. C.; Gawrisch, K.; Creamer, T. P. and White, S. H. (1996) *Proc. Natl. Acad. Sci. USA* 93, 2985-90.
- [174] Kumar, S. and Nussinov, R. (1999) *J. Mol. Biol.* 293, 1241-55.
- [175] Vijayakumar, M. and Zhou, H.-X. (2001) *J. Phys. Chem. B.* 105, 7334-40.
- [176] Kumar, S. and Nussinov, R. (2001) *Proteins* 43, 433-54.
- [177] Polticelli, F.; Ascenzi, P.; Bolognesi, M. and Honig, B. (1999) *Protein Sci.* 8, 2621-9.
- [178] Magalhaes, A.; Maigret, A.; Hoflack, J.; Gomes, J. N. and Scheraga, H. A. (1994) *J. Protein Chem.* 13, 195-215.
- [179] Vetriani, C.; Maeder, D. L.; Tolliday, N.; Yip, K. S.-P.; Stillman, T. J.; Britton, K. L.; Rice, D. W.; Klump, H. H. and Robb, F. T. (1998) *Proc. Natl. Acad. Sci. USA* 95, 12300-5.
- [180] Zhang, X.; Meining, W.; Fischer, M.; Bacher, A. and Ladenstein, R. (2001) *J. Mol. Biol.* 306, 1099-114.
- [181] Lurcu-Mustata, G.; VanBelle, D.; Wintjens, R.; Prevost, M. and Rooman, M. (2001) *Biopolymers* 59, 145-59.
- [182] Elcock, A. H. (1998) *J. Mol. Biol.* 284, 489-502.
- [183] Xiao, L. and Honig, B. (1999) *J. Mol. Biol.* 289, 1435-44.
- [184] Elcock, A. H. and McCammon, J. A. (1998) *J. Mol. Biol.* 280, 731-48.
- [185] Frolow, F.; Harel, M.; Sussman, J. L.; Mevarech, M. and Shoham, M. (1996) *Nat. Struct. Biol.* 3, 452-8.
- [186] Russel, R. J. M.; Gerike, U.; Danson, M. J.; Hough, D. W. and Taylor, G. L. (1998) *Structure* 6, 351-61.
- [187] Gorfe, A. A.; Brandsdal, B. O.; Leiros, H.-K. S.; Helland, R. and Smalås, A. O. (2000) *Proteins* 40, 207-17.
- [188] Tsai, C.-J.; Kumar, S.; Ma, B. and Nussinov, R. (1999) *Prot. Sci.* 8, 1181-90.
- [189] Tovchigrechko, A. and Vakser, I. A. (2001) *Protein Sci.* 10, 1572-83.
- [190] Tsai, C. J.; Ma, B.; Kumar, S.; Wolfson, H. and Nussinov, R. (2001) *Crit. Rev. Biochem. Mol. Biol.* 36, 399-433.
- [191] Ma, B.; Shatsky, M.; Wolfson, H. J. and Nussinov, R. (2002) *Protein Sci.* 11, 184-97.

- [192] Vaughan, C. K.; Buckle, A. M. and Fersht, A. R. (1999) *J. Mol. Biol.* 286, 1487-506.
- [193] Karplus, M. and Janin, J. (1999) *Prot. Eng.* 12, 185-6.
- [194] Privalov, P. L. and Tamura, A. (1999) *Prot. Eng.* 12, 187.
- [195] Pan, H.; Lee, C. and Hilser, V. J. (2000) *Proc. Natl. Acad. Sci. USA* 97, 12020-5.
- [196] Todd, M. J.; Luque, I.; Velazquez-Campoy, A. and Freire, E. (2000) *Biochemistry* 39, 11876-83.
- [197] Mian, I. S.; Bradwell, A. R. and Olson, A. J. (1991) *J. Mol. Biol.* 217, 133-51.
- [198] Kühlmann, U. C.; Pommer, A. J.; Moore, G. R.; James, R. and Kleanthous, C. (2000) *J. Mol. Biol.* 301, 1163-78.
- [199] Fersht, A. R. (1987) *TIBS* 12, 301-4.
- [200] Chothia, C. and Janin, J. (1975) *Nature* 256, 705-8.
- [201] Handschuh, S.; Goldfuss, B.; Chen, J.; Gasteiger, J. and Houk, K. N. (2000) *J. Comp-Aided. Mol. Design* 14, 611-29.
- [202] McCoy, A. J.; ChandanaEpa, V. and Colman, P. M. (1997) *J. Mol. Biol.* 268, 570-84.
- [203] Betts, M. J. and Sternberg, M. J. E. (1999) *Prot. Eng.* 12, 271-83.
- [204] Jackson, R. M. (1999) *Prot. Sci.* 8, 603-13.
- [205] DeLano, W. L.; Ultsch, M. H.; deVos, A. M. and Wells, J. A. (2000) *Science* 287, 1279-83.
- [206] Schreiber, G. and Fersht, A. R. (1996) *Nat. Struct. Biol.* 3, 427-31.
- [207] Selzer, T. and Schreiber, G. (2001) *Proteins* 45, 190-8.
- [208] Stites, W. E. (1997) *Chem. Rev.* 97, 1233-50.
- [209] Albeck, S.; Unger, R. and Schreiber, G. (2000) *J. Mol. Biol.* 298, 503-20.
- [210] Lum, K.; Chandler, D. and Weeks, J. D. (1999) *J. Phys. Chem. B* 103, 4570-7.
- [211] Huang, D. and Chandler, D. (2000) *Proc. Natl. Acad. Sci. USA* 97, 8324-7.
- [212] Elcock, A. H. and McCammon, J. A. (2001) *Proc. Natl. Acad. Sci. USA* 98, 2990-4.
- [213] Greenspan, N. S. and DiCera, E. (1999) *Nature Biotechnology* 17, 936-7.
- [214] Clackson, T. and Wells, J. A. (1995) *Science* 267, 383-6.
- [215] Bogan, A. A. and Thorn, K. S. (1998) *J. Mol. Biol.* 280, 1-9.
- [216] Young, L.; Jernigan, R. L. and Covell, D. G. (1994) *Protein Sci.* 3, 717-29.
- [217] Tsai, C.-J.; Lin, S. L.; Wolfson, H. J. and Nussinov, R. (1997) *Protein Sci.* 6, 53-64.
- [218] Hu, Z.; Ma, B.; Wolfson, H. and Nussinov, R. (2000) *Proteins* 39, 331-42.
- [219] Ma, B.; Wolfson, H. J. and Nussinov, R. (2001) *Curr. Opin. Struct. Biol.* 11, 364-9.
- [220] Valdar, W. S. J. and Thornton, J. M. (2001) *Proteins* 42, 108-24.
- [221] Lijnzaad, P.; Berendsen, H. J. C. and Argos, P. (1996) *Proteins* 25, 389-97.
- [222] Jones, S.; Marin, A. and Thornton, J. M. (2000) *Prot. Eng.* 13, 77-82.
- [223] LoConte, L.; Chothia, C. and Janin, J. (1999) *J. Mol. Biol.* 285, 2177-98.
- [224] Zhu, Z.-Y. and Karlin, S. (1996) *Proc. Natl. Acad. Sci. USA* 93, 8350-5.
- [225] Sheinerman, F. B.; Norel, R. and Honig, B. (2000) *Curr. Opin. Struct. Biol.* 10, 153-9.
- [226] Covell, D. G. and Wallqvist, A. (1997) *J. Mol. Biol.* 269, 281-97.
- [227] Camacho, C. J.; Kimura, S. R.; DeLisi, C. and Vadja, S. (2000) *Biophys. J.* 78, 1094-105.
- [228] Zhang, C.; Chen, J. and DeLisi, C. (1999) *Proteins* 34, 255-67.
- [229] McCammon, J. A. (1998) *Curr. Opin. Struct. Biol.* 8, 245-9.
- [230] Kollman, P. A. (1994) *Curr. Opin. Struct. Biol.* 4, 240-5.
- [231] Murray-Rust, P. (1994) *Curr. Opin. Biotechnol.* 5, 648-53.
- [232] Gane, P. J. and Dean, P. M. (2000) *Curr. Opin. Struct. Biol.* 10, 401-4.
- [233] Stahl, M. and Rarey, M. (2001) *J. Med. Chem.* 44, 1035-42.
- [234] Shoichet, B. K. and Kuntz, I. D. (1991) *J. Mol. Biol.* 221, 327-46.
- [235] Jackson, R. M. and Sternberg, M. J. E. (1995) *J. Mol. Biol.* 250, 258-75.
- [236] Morelli, X.; Czjzek, M.; Hatchikian, C. E.; Bornet, O.; Fontecilla-Camps, J. C.; Palma, N. P.; Moura, J. J. and Guerlesquin, F. (2000) *J. Biol. Chem.* 275, 23204-10.
- [237] Böhm, H. J. (1998) *J. Comp-Aided Mol. Design* 12, 1-15.
- [238] Shoichet, B. K.; Leach, A. R. and Kuntz, I. D. (1999) *Proteins* 34, 4-16.
- [239] Gohlke, H. and Klebe, G. (2001) *Curr. Opin. Struct. Biol.* 11, 231-5.
- [240] Hansson, T. and Aqvist, J. (1995) *Prot. Eng.* 8, 1137-44.
- [241] Morelli, X. J.; Palma, P. N.; Guerlesquin, F. and Rigby, A. C. (2001) *Protein Sci.* 10, 2131-7.
- [242] Norel, R.; Sheinerman, F.; Petrey, D. and Honig, B. (2001) *Protein Sci.* 10, 2147-61.
- [243] Zhou, R.; Friesner, R. A.; Ghosh, A.; Rizzo, R. C.; Jorgensen, W. L. and Levy, R. M. (2001) *J. Phys. Chem. B* 105, 10388-97.

- [244] Noskov, S. Y. and Lim, C. (2001) *Biophys. J.* 81, 737-50.
- [245] Fernandez-Recio, J.; Totrov, M. and Abagyan, R. (2002) *Protein Sci.* 11, 280-91.
- [246] Liang, J.; Edelsbrunner, H. and Woodward, C. (1998) *Protein Sci.* 7, 1884-97.
- [247] Verdonk, M. L.; Cole, J. C. and Taylor, R. (1999) *J. Mol. Biol.* 289, 1093-108.
- [248] Gallet, X.; Charlotheaux, B.; Thomas, A. and Brasseur, R. (2000) *J. Mol. Biol.* 302, 917-26.
- [249] Lichtarge, O.; Bourne, H. R. and Cohen, F. E. (1996) *J. Mol. Biol.* 257, 342-58.
- [250] Armon, A.; Graur, D. and Ben-Tal, N. (2001) *J. Mol. Biol.* 307, 447-63.
- [251] Kuntz, I. D. (1992) *Science* 257, 1078-82.
- [252] Miranker, A. and Karplus, M. (1991) *Proteins* 11, 29-34.
- [253] Pazos, F.; Helmer-Citterich, M.; Ausiello, G. and Valencia, A. (1997) *J. Mol. Biol.* 271, 511-23.
- [254] Russel, R. B.; Sasieni, P. D. and Sternberg, M. J. E. (1998) *J. Mol. Biol.* 282, 903-18.
- [255] Schmitt, S.; Hendlich, M. and Klebe, G. (2001) *Angew. Chem. Int. Ed.* 40, 3141-4.
- [256] Aloy, P.; Querol, E.; Aviles, F. X. and Sternberg, M. J. E. (2001) *J. Mol. Biol.* 311, 395-408.
- [257] Jiang, L.; Gao, Y.; Mao, F.; Liu, Z. and Lai, L. (2002) *Proteins* 46, 190-6.
- [258] Gilson, M. K.; Given, J. A.; Bush, B. L. and McCammon, J. A. (1997) *Biophys. J.* 72, 1047-69.
- [259] Elcock, A. H.; Sept, D. and McCammon, J. A. (2001) *J. Phys. Chem. B.* 105, 1504-18.
- [260] Barril, X.; Gelpi, J. L.; Lopez, J. M.; Orozco, M. and Luque, F. J. (2001) *Theor. Chem. Acc.* 106, 2-9.
- [261] Qasba, P. K. (2000) *Carbohydrate Polymers* 41, 293-309.
- [262] Essmann, U. and Berkowitz, M. L. (1999) *Biophys. J.* 76, 2081-9.
- [263] Tieleman, D. P.; Marrink, S. J. and Berendsen, H. J. C. (1997) *Biochim. Biophys. Acta* 1331, 235-70.
- [264] Kessel, A.; Cafiso, D. S. and Ben-Tal, N. (2000) *Biophys. J.* 78, 571-83.
- [265] Yau, W.-M.; Wimley, W. C.; Gawrisch, K. and White, S. H. (1998) *Biochemistry* 37, 14713-8.
- [266] Zwaal, R. F. A.; Comfurius, P. and Bevers, E. M. (1998) *Biochim. Biophys. Acta* 1376, 433-53.
- [267] Nelsestuen, G. L. and Ostrowski, B. G. (1999) *Curr. Opin. Struct. Biol.* 9, 433-7.
- [268] Tobias, D. J. (2001) *Curr. Opin. Struct. Biol.* 11, 253-61.
- [269] Murray, D.; Ben-Tal, N.; Honig, B. and McLaughlin, S. (1997) *Structure* 5, 985-9.
- [270] Juffer, A. H.; Shepherd, C. M. and Vogel, H. J. (2001) *J. Chem. Phys.* 114, 1892-905.
- [271] Bursavich, M. G.; West, C. W. and Rich, D. H. (2001) *Org. Lett.* 3, 2317-20.
- [272] Greenspan, N. S. (2001) *Adv. Cancer Research* 80, 147-87.
- [273] Perrett, S.; Zahn, R.; Stenberg, G. and Fersht, A. R. (1997) *J. Mol. Biol.* 269, 892-901.
- [274] Friedrich, U.; Blom, A. M.; Dahlback, B. and Villoutreix, B. O. (2001) *J. Biol. Chem.* 276, 24122-8.
- [275] Olson, S. T.; Halvorson, H. R. and Bjork, I. (1991) *J. Biol. Chem.* 266, 6342-52.
- [276] Bross, P.; Corydon, T. J.; Andresen, B. S.; Jorgensen, M. M.; Bolund, L. and Gregersen, N. (1999) *Hum. Mut.* 14, 186-98.
- [277] Wickner, S.; Maurizi, M. R. and Gottesman, S. (1999) *Science* 286, 1888-93.
- [278] Weatherall, D. J. (2001) *Nature Genetics* 2, 245-55.
- [279] Carrell, R. W. and Stein, P. E. (1996) *Biol. Chem. Hoppe-Seyler* 377, 1-17.
- [280] Carrell, R. W. and Lomas, D. A. (2002) *N. Engl. J. Med.* 346, 45-53.
- [281] Taverna, D. M. and Goldstein, R. A. (2002) *Proteins* 46, 105-9.
- [282] Taverna, D. M. and Goldstein, R. A. (2002) *J. Mol. Biol.* 315, 479-84.
- [283] Feldman, D. E. and Frydman, J. (2000) *Curr. Opin. Struct. Biol.* 10, 26-33.
- [284] Zhou, H.-X. and Dill, K. A. (2001) *Biochemistry* 40, 11289-93.
- [285] Sinha, N. and Nussinov, R. (2001) *Proc. Natl. Acad. Sci. USA* 98, 3139-44.
- [286] Sunyaev, S.; III, W. L. and Bork, P. (2001) *Curr. Opin. Struct. Biol.* 11, 125-30.
- [287] Wang, Z. and Moulton, J. (2001) *Hum. Mut.* 17, 263-70.
- [288] Karpusas, M.; Baase, W. A.; Matsumura, M. and Matthews, B. W. (1989) *Proc. Natl. Acad. Sci. USA* 86, 8237-41.
- [289] Bowie, J. U.; Reidhaar-Olson, J. F.; Lim, W. A. and Sauer, R. T. (1990) *Science* 247, 1306-10.
- [290] Lim, W. A.; Farruggio, D. C. and Sauer, R. T. (1992) *Biochemistry* 31, 4324-33.
- [291] Freire, E. (1999) *Proc. Natl. Acad. Sci. USA* 96, 10118-22.
- [292] Durand, P.; Fabrega, S.; Henrissat, B.; Mornon, J.-P. and Lehn, P. (2000) *Hum. Mol. Genetics* 9, 967-77.
- [293] Covalt, J. C.; Roy, M. and Jennings, P. A. (2001) *J. Mol. Biol.* 307, 657-69.
- [294] Kellis, J. T.; Nyberg, K.; Sali, D. and Fersht, A. R. (1988) *Nature* 333, 784-6.

- [295] Xu, B.; Hua, Q.-X.; Nakagawa, S. H.; Jia, W.; Chu, Y.-C.; Katsoyannis, P. G. and Weiss, M. A. (2002) *Protein Sci.* 11, 104-16.
- [296] Cota, E.; Hamill, S. J.; Fower, S. B. and Clarke, J. (2000) *J. Mol. Biol.* 302, 713-25.
- [297] Chakravarty, S. and R, R. V. (1999) *Structure Fold. Des.* 7, 723-32.
- [298] Axe, D. D. (2000) *J. Mol. Biol.* 301, 585-95.
- [299] Luque, I. and Freire, E. (2000) *Proteins Suppl* 4, 63-71.
- [300] Pace, C. N.; Alston, R. W. and Shaw, K. L. (2000) *Protein Sci.* 9, 1395-8.
- [301] Anderson, D. E.; Becktel, W. J. and Dahlquist, F. W. (1990) *Biochemistry* 29, 2403-8.
- [302] Chen, Y. W.; Fersht, A. R. and Henrick, K. (1993) *J. Mol. Biol.* 234, 1158-70.
- [303] Kumar, S.; Tsai, C.-J. and Nussinov, R. (2000) *Prot. Eng.* 13, 179-91.
- [304] Marqusee, S. and Sauer, R. T. (1994) *Protein Sci.* 3, 2217-25.
- [305] Matthews, J. M.; Norton, R. S.; Hammacher, A. and Simpson, R. J. (2000) *Biochemistry* 39, 1942-50.
- [306] Cordes, M. H. J. and Sauer, R. T. (1999) *Protein Sci.* 8, 318-25.
- [307] Perl, D.; Mueller, U.; Heinemann, U. and Schmid, F. X. (2000) *Nat. Struct. Biol.* 7, 380-3.
- [308] Matthews, B. W.; Nicholson, H. and Becktel, W. J. (1987) *Proc. Natl. Acad. Sci. USA* 84, 6663-7.
- [309] Mok, Y.-K.; Elisseeva, E. L.; Davidson, A. R. and Formankay, J. D. (2001) *J. Mol. Biol.* 307, 913-28.
- [310] Massi, F. and Straub, J. E. (2001) *Biophys. J.* 81, 697-709.
- [311] Word, J. M.; Bateman, R. C.; Presley, B. K.; Lovell, S. C. and Richardson, D. C. (2000) *Protein Sci.* 9, 2251-9.
- [312] Gillis, D. and Rومان, M. (2000) *Prot. Eng.* 13, 849-56.
- [313] Mendes, J.; Baptista, A. M.; Carrondo, M. A. and Soares, C. M. (2001) *J. Comp-Aided Mol. Design* 15, 721-40.
- [314] DeFilippis, V.; Sander, C. and Vriend, G. (1994) *Protein Eng.* 10, 1203-8.
- [315] Dahlback, B. (2000) *The Lancet* 355, 1627-32.
- [316] Daura, X.; Haaksma, E. and vanGunsteren, W. F. (2000) *J. Comp-Aided Mol. Design* 14, 507-29.
- [317] Perera, L.; Darden, T. A. and Pedersen, L. G. (2001) *Thromb. Haemost.* 85, 596-603.
- [318] Wiley, M. R.; Weir, L. C.; Briggs, S.; Bryan, N. A.; Buben, J.; Campbell, C.; Chirgadze, N. Y.; Conrad, R. C.; Craft, T. J.; Ficorilli, J. V.; Franciskovich, J. B.; Froelich, L. L.; Gifford-Moore, D. S.; Goodson, T. J.; Herron, D. K.; Klimkowski, V. J.; Kyle, K. D.; Masters, J. J.; Ratz, A. M.; Milot, G.; Shuman, R. T.; Smith, T.; Smith, G. F.; Tebbe, A. L. and Tinsley, J. M. (2000) *J. Med. Chem.* 43, 883-99.
- [319] Pierce, A. C. and Jorgensen, W. L. (2001) *J. Med. Chem.* 44, 1043-50.
- [320] Dullweber, F.; Stubbs, M. T.; Musil, D.; Sturzebecher, J. and Klebe, G. (2001) *J. Mol. Biol.* 313, 593-614.
- [321] Badellino, K. O. and Walsh, P. N. (2001) *Biochemistry* 40, 7569-80.
- [322] Henriques, E. S.; Floriano, W. B.; Reuter, N.; Melo, A.; Brown, D.; Gomes, J. A. N. F.; Maignet, B.; Nascimento, M. A. C. and Ramos, M. J. (2001) *J. Comput. Aided Mol. Des.* 15, 309-22.
- [323] Villoutreix, B. O. and Dahlbäck, B. (1998) *J. Mol. Mod.* 4, 310-22.
- [324] Light, D. R.; Glaser, C. B.; Betts, M.; Blasko, E.; Campbell, E.; Clarke, J. H.; McCaman, M.; McLean, K.; Nagashima, M.; Parkinson, J. F.; Rumennik, G.; Young, T. and Morser, J. (1999) *Eur. J. Biochem.* 262, 522-33.
- [325] Bitomsky, W. and Wade, R. C. (1999) *J. Am. Chem. Soc.* 121, 3004-13.
- [326] Faioni, E. M.; Hermida, J.; Rovida, E.; Razzari, C.; Asti, D.; Zeinali, S. and Mannucci, P. M. (2000) *Br. J. Haematol.* 108, 265-71.
- [327] Villoutreix, B. O.; Åkerstrom, B. and Lindqvist, A. (2000) *Blood Coagul Fibrinolysis* 11, 261-76.
- [328] Deam, S.; Srinivasan, N.; Westby, J.; Horn, E. H. and Dolan, G. (2001) *Thromb. Haemost.* 85, 265-9.
- [329] Ichinose, A.; Tsukamoto, H.; Izumi, T.; Yamazaki, T.; Togashi, M.; Takamatsu, J.; Saito, H. and Umeyama, H. (1998) *Br. J. Haematol.* 101, 264-72.
- [330] Akhavan, S.; Mannucci, P. M.; Lak, M.; Mancuso, G.; Mazzucconi, M. G.; Rocino, A.; Jenkins, P. V. and Perkins, S. J. (2000) *Thromb. Haemost.* 84, 989-97.
- [331] Ashton, A. W.; Boehm, M. K.; Johnson, D. J. D.; Kembal-Cook, G. and Perkins, S. J. (1998) *Biochemistry* 37, 8208-17.
- [332] Rester, U.; Bode, W.; Moser, M.; Parry, M. A. A.; Huber, R. and Auerswald, E. (1999) *J. Mol. Biol.* 293, 93-106.
- [333] Jespers, L.; Vanwetswinkel, S.; Lijnen, H. R.; Herzelee, N. V.; Hoef, B. V.; Demarsin, E.; Collen, D. and Maeyer, M. D. (1999) *Thromb. Haemost.* 81, 479-85.
- [334] Manfioletti, G.; Brancolini, C.; Avanzi, G. and Schneider, C. (1993) *Mol. Cell. Biol.* 13, 4976-85.
- [335] Gandrille, S.; Borgel, D.; Sala, N.; Espinosa-Parrilla, Y.; Simmonds, R.; Rezende, S.; Lind, B.; Mannhalter, C.; Pabinger, I.; Reitsma, P. H.; Formstone, C.; Cooper, D. N.; Saito, H.; Suzuki, K.; Bernardi, F. and Aiach, M. (2000) *Thromb. Haemost.* 84, 918.
- [336] Stenflo, J. (1999) *Crit. Rev. Eukaryotic Gene Expression* 9, 59-88.
- [337] Dahlbäck, B. (1991) *Thromb. Haemost.* 66, 49-58.
- [338] Beck, K.; Hunter, I. and Engel, J. (1990) *FASEB J.* 4, 148-60.
- [339] Joseph, D. R. and Baker, M. E. (1992) *FASEB J.* 6, 2477-81.

- [340] Hillarp, A. and Dahlbäck, B. (1988) *J. Biol. Chem.* 263, 12759-64.
- [341] Villoutreix, B. O.; Blom, A. M.; Webb, J. and Dahlbäck, B. (1999) *Immunopharmacology* 42, 121-34.
- [342] Fernández, J. A. and Griffin, J. H. (1994) *J. Biol. Chem.* 269, 2535-40.
- [343] Fernández, J. A.; Villoutreix, B. O.; Hackeng, T. M.; Griffin, J. H. and Bouma, B. N. (1994) *Biochemistry* 33, 11073-8.
- [344] Härdig, Y. and Dahlbäck, B. (1996) *J. Biol. Chem.* 271, 20861-7.
- [345] Villoutreix, B. O.; Teleman, O. and Dahlbäck, B. (1997) *J. Comp-Aided Mol. Design* 11, 293-304.
- [346] Giri, T. K.; Villoutreix, B. O.; Wallqvist, A.; Dahlbäck, B. and deFrutos, P. G. (1998) *Thromb. Haemost.* 80, 798-804.
- [347] He, X.; Shen, L.; Villoutreix, B. O. and Dahlbäck, B. (1998) *J. Biol. Chem.* 273, 27449-58.
- [348] Giri, T. K.; deFrutos, P. G.; Yamazaki, T.; Villoutreix, B. O. and Dahlbäck, B. (1999) *Thromb. Haemost.* 82, 1627-33.
- [349] Soriano-Garcia, M.; Padmanabhan, K.; deVos, A. M. and Tulinsky, A. (1992) *Biochemistry* 31, 2554-66.
- [350] Sunnerhagen, M.; Forsen, S.; Hoffren, A. M.; Drakenberg, T.; Teleman, O. and Stenflo, J. (1995) *Nat. Struct. Biol.* 2, 504-9.
- [351] Borgel, D.; Gaussem, P.; Garbay, C.; Bachelot-Loza, C.; Kaabache, T.; Liu, W.-Q.; Brohard-Bohn, B.; LeBonniec, B.; Aiach, M. and Gandrille, S. (2001) *Biochem. J.* 360, 499-506.
- [352] Gandrille, S.; Borgel, D.; Eschwege-Gufflet, V.; Aillaud, M.; Dreyfus, M.; Matheron, C.; Gaussem, P.; Abgrall, J. F.; Jude, B.; Sie, P.; Toulon, P. and Aiach, M. (1995) *Blood* 85, 130-8.
- [353] Villoutreix, B. O.; deFrutos, P. G.; Lovenklev, M.; Linse, S.; Fernlund, P. and Dahlbäck, B. (1997) *Proteins* 29, 478-91.
- [354] Beckmann, G.; Hanke, J.; Bork, P. and Reich, J. G. (1998) *J. Mol. Biol.* 275, 725-30.
- [355] Grishkovskaya, I.; Avvakumov, G. V.; Sklenar, G.; Dales, D.; Hammond, G. L. and Muller, Y. A. (2000) *EMBO J.* 19, 504-12.
- [356] Hohenester, E.; Tisi, D.; Talts, J. F. and Timpl, R. (1999) *Mol. Cell.* 4, 783-92.
- [357] Tisi, D.; Talts, J. F.; Timpl, R. and Hohenester, E. (2000) *EMBO J.* 19, 1432-40.
- [358] Rudenko, G.; Nguyen, T.; Chelliah, Y.; Sudhof, T. C. and Deisenhofer, J. (1999) *Cell* 99, 93-101.
- [359] Villoutreix, B. O.; Dahlbäck, B.; Borgel, D.; Gandrille, S. and Muller, Y. A. (2001) *Proteins* 43, 203-16.
- [360] Andersen, B. D.; Bisgaard, M. L.; Lind, B.; Philips, M.; Villoutreix, B. O. and Thorsen, S. (2001) *Thromb. Haemost.* 86, 1392-9.
- [361] Villoutreix, B. O.; Fernández, J. A.; Teleman, O. and Griffin, J. H. (1995) *Prot. Eng.* 8, 1253-8.
- [362] Blom, A. M.; Covell, D. G.; Wallqvist, A.; Dahlbäck, B. and Villoutreix, B. O. (1998) *Biochim. Biophys. Acta* 1388, 181-9.
- [363] Webb, J.; Villoutreix, B. O.; Dahlback, B. and Blom, A. M. (2001) *J. Biol. Chem.* 276, 4330-7.
- [364] Villoutreix, B. O.; Härdig, Y.; Wallqvist, A.; Covell, D. G.; Frutos, P. G. d. and Dahlbäck, B. (1998) *Proteins* 31, 391-405.
- [365] Blom, A. M.; Webb, J.; Villoutreix, B. O. and Dahlbäck, B. (1999) *J. Biol. Chem.* 274, 19237-45.
- [366] Bode, W. and Huber, R. (2000) *Biochim. Biophys. Acta* 1477, 241-52.
- [367] Mather, T.; Oganessyan, V.; Hof, P.; Huber, R.; Foundling, S.; Esmon, C. and Bode, W. (1996) *EMBO J.* 15, 6822-31.
- [368] Fisher, C. L.; Greengard, J. S. and Griffin, J. H. (1994) *Protein Sci.* 3, 588-99.
- [369] Wacey, A. I.; Pemberton, S.; Cooper, D. N.; Kakkar, V. V. and Tuddenham, E. G. D. (1993) *Br. J. Haemato.* 84, 290-300.
- [370] Greengard, J. S.; Fisher, C. L.; Villoutreix, B. and Griffin, J. H. (1994) *Proteins* 18, 367-80.
- [371] Wacey, A. I.; Krawczak, M.; Kembell-Cook, G. and Cooper, D. N. (1997) *Hum. Genet.* 101, 37-42.
- [372] Perera, L.; Foley, C.; Darden, T. A.; Stafford, D.; Mather, T.; Esmon, C. T. and Pedersen, L. G. (2000) *Biophys. J.* 79, 2925-43.
- [373] Villoutreix, B. O.; Covell, D. G.; Blom, A. M.; Wallqvist, A.; Friedrich, U. and Dahlbäck, B. (2001) *J. Comput. Aided. Mol. Des.* 15, 13-27.
- [374] Knobe, K. E.; Berntsdotter, A.; Shen, L.; Morser, J.; Dahlbäck, B. and Villoutreix, B. O. (1999) *Proteins* 35, 218-34.
- [375] Esmon, C. T. (2000) *Thromb. Haemost.* 83, 639-43.
- [376] Villoutreix, B. O.; Blom, A. M. and Dahlbäck, B. (1999) *Prot. Eng.* 12, 833-40.
- [377] Simmonds, R. E. and Lane, D. A. (1999) *Blood* 94, 632-41.
- [378] Liaw, P. C.; Mather, T.; Oganessian, N.; Ferrell, G. L. and Esmon, C. T. (2001) *J. Biol. Chem.* 276, 8364-70.
- [379] Kalafatis, M.; Egan, J. O.; Veer, C. v. t.; Cawthern, K. M. and Mann, K. G. (1997) *Crit. Rev. Eukaryot. Gene Expr.* 7, 241-80.
- [380] Lenting, P. J.; Mourik, J. A. v. and Mertens, K. (1998) *Blood* 92, 3983-96.
- [381] Rosing, J. and Tans, G. (1997) *Thromb. Haemost.* 78, 427-33.
- [382] Murphy, M. E. P.; Lindley, P. F. and Adman, E. T. (1997) *Protein Sci.* 6, 761-70.

- [383] Pemberton, S.; Lindley, P.; Zaitsev, V.; Card, G.; Tuddenham, E. G. D. and Kemball-Cook, G. (1997) *Blood* 89, 2413-21.
- [384] Pan, Y.; DeFay, T.; Gitschier, J. and Cohen, F. E. (1995) *Nat. Struct. Biol.* 2, 740-4.
- [385] Villoutreix, B. O. and Dahlbäck, B. (1998) *Protein Sci.* 7, 1317-25.
- [386] Pellequer, J.; Gale, A. J.; Getzoff, E. D. and Griffin, J. H. (2000) *Thromb. Haemost.* 84, 849-57.
- [387] Zaitseva, I.; Zaitsev, V.; Card, G.; Moshkov, K.; Bax, B.; Ralph, A. and Lindley, P. (1996) *J. Biol. Inorg. Chem.* 1, 15-23.
- [388] Stoylova, S. S.; Lenting, P. J.; Kemball-Cook, G. and Holzenburg, A. (1999) *J. Biol. Chem.* 274, 36573-8.
- [389] Castoldi, E.; Simioni, P.; Kalafatis, M.; Lunghi, B.; Tormene, D.; Girelli, D.; Girolami, A. and Bernardi, F. (2000) *Blood* 96, 1443-8.
- [390] vanWijk, R.; Nieuwenhuis, K.; Berg, M. v. d.; Huizinga, E. G.; Meijden, B. v. d.; Kraaijenhagen, R. J. and vanSolinge, W. W. (2001) *Blood* 98, 358-67.
- [391] Thompson, A. R.; Murphy, M. E. P.; Liu, M.; Saenko, E. L.; Healey, J. F.; Lollar, P. and Scandella, D. (1997) *Blood* 90, 1902-10.
- [392] Schwaab, R.; Oldenburg, J.; Lalloz, M. R. A.; Schwaab, U.; Pemberton, S.; Hanfland, P.; Brackmann, H. H.; Tuddenham, E. G. D. and Michaelides, K. (1997) *Hum. Genet.* 101, 323-32.
- [393] Kemball-Cook, G.; Tuddenham, E. G. D. and Wacey, A. I. (1998) *Nucleic Acids Res.* 26, 216-9.
- [394] Amano, K.; Sarkar, R.; Pemberton, S.; Kemball-Cook, G.; Kazazian, H. H. and Kaufman, R. J. (1998) *Blood* 91, 538-48.
- [395] Bajaj, S. P.; Schmidt, A. E.; Mathur, A.; Padmanabhan, K.; Zhong, D.; Mastro, M. and Fay, P. J. (2001) *J. Biol. Chem.* 276, 16302-9.
- [396] Vogel, W. (1999) *FASEB J.* 13, S77-S82.
- [397] Pratt, K. P.; Shen, B. W.; Takeshima, K.; Davie, E. W.; Fujikawa, K. and Stoddard, B. L. (1999) *Nature* 402, 439-42.
- [398] Macedo-Ribeiro, S.; Bode, W.; Huber, R.; Quinn-Allen, M. A.; Kim, S. W.; Ortel, T. L.; Bourenkov, G. P.; Bartunik, H. D.; Stubbs, M. T.; Kane, W. H. and Fuentes-Prior, P. (1999) *Nature* 402, 434-9.
- [399] Ito, N.; Phillips, S. E. V.; Yadav, K. D. S. and Knowles, P. F. (1994) *J. Mol. Biol.* 238, 794-814.
- [400] Fischer, D. and Eisenberg, D. (1996) *Protein Sci.* 5, 947-55.
- [401] Villoutreix, B. O.; Bucher, P.; Hofmann, K.; Baumgartner, S. and Dahlbäck, B. (1998) *J. Mol. Mod.* 4, 268-75.
- [402] Pellequer, J.-L.; Gale, A. J.; Griffin, J. H. and Getzoff, E. D. (1998) *Blood Cells Mol. Dis.* 24, 448-61.
- [403] Gale, A. J.; Pellequer, J.-L.; Getzoff, E. D. and Griffin, J. H. (2000) *Thromb. Haemost.* 83, 78-85.
- [404] Gilbert, G. E. and Baleja, J. D. (1995) *Biochemistry* 34, 3022-31.
- [405] Nicolaes, G. A. F.; Villoutreix, B. O. and Dahlbäck, B. (1999) *Biochemistry* 38, 13584-91.
- [406] Nicolaes, G. A. F.; Villoutreix, B. O. and Dahlbäck, B. (2000) *Blood Coagul. Fibrinolysis* 11, 89-100.
- [407] Knobe, K. E.; Villoutreix, B. O.; Tengborn, L. I.; Petrini, P. and Ljung, R. C. R. (2000) *Haemostasis* 30, 268-79.
- [408] Liu, M.-L.; Shen, B. W.; Nakaya, S.; Pratt, K. P.; Fujikawa, K.; Davie, E. W.; Stoddard, B. L. and Thompson, A. R. (2000) *Blood* 96, 979-87.
- [409] Pratt, K. P. (2000) *Curr. Opin. Drug Discov. & Development* 3, 516-26.
- [410] Grimm, D. R.; Colter, M. B.; Braunschweig, M.; Alexander, L. J.; Neame, P. J. and Kim, H. K. W. (2001) *Cell Mol. Life. Sci.* 58, 148-59.
- [411] Stoylova-McPhie, S.; Villoutreix, B. O.; Mertens, K.; Kemball-Cook, G. and Holzenburg, A. (2002) *Blood* 99, 1215-23.