Computational Investigation of Selected Protein Families

Thesis submitted for the degree
"Doctor of Philosophy"

By Meytal Landau

Submitted to the Senate of Tel-Aviv University
December, 2007

The work was carried out under the supervision of **Prof. Nir Ben-Tal**

Acknowledgements

My deepest gratitude goes to Prof. Nir Ben-Tal, who has been much more than an admirable supervisor. Thank you for your empathy, patience, generosity and your willingness to share your remarkable knowledge and wisdom. I most appreciate the way you always encouraged me to follow my own visions.

Many thanks to my lab members: Adva Suez, Dalit Shental, Fabian Glaser, Gilad Wainreb, Guy Nimrod, Maya Schushan, Uri Ron, Yanay Ofran, Yariv Barkan, additional and former members of the Ben-Tal group, for the wonderful experience of working together.
Special thanks to Sarel Fleishman for his enormous help, priceless insights and great friendship.

I am grateful for having the opportunity to work with Prof. Uri Seligsohn, Dr. Ariella Zivelin, Dr. Nurit Rosenberg and the other "coagulation people". It has been tremendously fruitful and also a lot of fun.
I also appreciate the invaluable and enjoyable collaboration with Prof. Etana Padan (The Hebrew University).

I wish to express my gratitude and love to my family: my brilliant husband, Arik, for being who he is, my dear parents, Uri and Edith, for their love and support, my sister Sarai for her deep observations, and my family-in-law, Jenoy, Ricu, Henit, Avner, Roy and Noa for the warm welcome into their lives.

This work is dedicated with

love to my life partner Arik

and my son Daniel.

Table of Contents

SYNOPSIS

The main objective of my doctoral research has been to increase our understanding of the molecular details of the regulatory mechanisms of proteins that play key roles in physiological processes and diseases. To achieve that goal, I focused on protein structures, and integrated data from biochemical and clinical studies with computational analyses that I performed on these proteins based on their evolutionary history and physicochemical properties. Specifically, I pursued the mechanistic understanding of the function and regulation of two transmembrane protein families, namely the ErbB family and Na^+/H^+ exchangers. Both families are involved in devastating human diseases and advancing our knowledge of their activation is of high clinical importance.

My earliest interest was in shedding some light on the molecular regulation of the ErbB family. Tyrosine kinase receptors of the ErbB family play a significant role in vital cellular processes and in various cancers. These receptors are unique among kinases in that their activation is independent of their phosphorylation state. Moreover, the first crystal structure of the kinase domain of ErbB1, the first member to be discovered that is also known as the epidermal growth factor receptor (EGFR), showed that all elements are ready for catalysis in the un-phosphorylated state. Thus, the available data raised an apparent paradox, since it was well known that the EGFR is not constitutively active. Consequently, I had sought an alternative regulation mechanism. My working hypothesis was that these receptors are regulated by another mechanism intrinsic to the intracellular domain, which is phosphorylation-independent. My results indicated that the C-terminal domain of the EGFR serves as a down-regulator of the catalytic activity via direct contacts with the kinase domain. Recent publications substantiated my findings by experimental means.

Since the publication of my results, abundant data, including structural and biochemical studies, has been accumulated. Base of this information, I recently summarized a model that includes multiple layers of regulation on EGFR activity imposed via its different structural domains, and their mutual control. The model provides an explanation, at the molecular level, for the effects of cancer-causing ErbB mutations, and suggests a novel therapeutic venue for ErbB-related cancer.

Similar to the ErbB family, the Na^+/H^+ exchangers raise interesting questions regarding structure-function relationships within transmembrane proteins. These proteins transport Na^+

and H^+ ions across the membrane and play a key role in maintaining cellular homeostasis and in pathological conditions such as cancer and heart diseases. Despite extensive mutagenesis studies, the molecular details of the cation-exchange mechanism are still obscure, largely because the structure of these transporters is unknown. However, due to various technical difficulties with the expression and purification of these proteins, it appears that we are still years away from structure determination.

The structure of the Na^+/H^+ antiporter from *Escherichia coli* (EcNhaA) was recently determined using X-ray crystallography. This unique structure opened the way for understanding structure-function relation in human transporters. EcNhaA and eukaryotic Na^+/H^+ exchangers appear to share similar functions and architecture, and I therefore used that crystal structure to model the structure of the human Na^+/H^+ exchanger isoform 1 (NHE1). Modeling was extremely challenging because the sequence identity of these proteins is about 10%, well below the cutoff point for reliable comparative modeling. I therefore combined the outcome of several fold-recognition and sequence-alignment tools with manual interventions. My model structure was validated according to various criteria, such as the evolutionary conservation profile and mutagenesis data.

The model structure is especially valuable when used to guide biochemical experiments. Accordingly, I established collaboration with Prof. Etana Padan (The Hebrew University), who was involved in the determination of the EcNhaA structure. Specifically, in my model, amino acids that participate in the binding of clinically-important NHE1-inhibitors are found in close proximity to each other in the structure, even though they are scattered on the primary sequence. This provided important support to the model and also allowed me to infer the location of the binding-site of inhibitors within the EcNhaA structure. The prediction was validated by site-directed mutagenesis done by Katia Herz under the supervision of Etana Padan. This result strongly supported my motivation to predict the structure of NHE1 on the basis of the crystal structure of the prokaryotic protein.

I further point out that the two transporters share a similar cation-exchange mechanism. This allowed me to infer mechanistic insights available for the prokaryotic protein in order to further our understanding of the mechanism of NHE1 activity. Thus, I integrated empirical data and the new structural model to suggest an alternating-access mechanism of the Na^+/H^+ exchange in molecular detail.

INTRODUCTION

My main drive in research has been the investigation of function and regulatory mechanisms of proteins in molecular details. Such information is imperative when trying to accurately control and intervene with cellular processes, e.g., in pathological conditions. Understanding protein mechanisms at the molecular level require structural information, such as via X-ray crystallography [1]. Nevertheless, obtaining 3-dimensional (3D) structures is not enough, mainly since they typically display a static view, while proteins are dynamic entities, and conformational changes are a part of their regulation and activity [2]. Consequently, my research approach was to exploit available 3D-structures or to computationally predict them, and to complement the structural information with sequence, phylogenetic and biophysical analyses in order to better understand their structure-function relationships. I then integrated the computational information with available biochemical and genetic data to shed light on the regulatory mechanisms. Specifically, my doctoral dissertation focused on the investigation of two transmembrane protein families, namely the ErbB family (described in chapter 1) and the Na^+/H^+ exchanger family (described in chapter 2).

My particular research makes use of the available algorithms and biological databases in order to address biological problems and to guide hypothesis-driven experiments. A primary approach that I utilized in my investigation of protein families was the examination of their evolutionary conservation. The main idea is that key amino acids that are important for the maintenance of the structure and function of a protein evolve slowly, i.e., are more evolutionary conserved than unessential residues. The ConSurf server (http://consurf.tau.ac.il/) is a web-based tool developed in the laboratory of Prof. Nir Ben-Tal in collaboration with Dr. Tal Pupko (Tel-Aviv University) [3]. ConSurf calculates conservation scores for each amino-acid site, based on the phylogenetic relations between homologous sequences as well as by a specified stochastic process of amino-acid replacements [4]. The evolutionary conservation scores are mapped onto the 3D-structure of proteins, which are visualized using a user-friendly interface. I was involved in the integration of new features into ConSurf to further improve the accuracy of the calculations, which has lead to the publication of ConSurf version 3.0 [5] (Landau et. al. (2005) *Nucleic Acids Research* 33, W299-302; the manuscript is attached as Appendix I). My role in this project was to correlate between computation and biology. I examined the implications of modifications in the computational methods and algorithms implemented in ConSurf on the analyses of biological systems. In addition, I was responsible for setting the parameters used

in these algorithms that would best describe the evolutionary conservation profile of protein families.

Evolutionary conservation analyses provide valuable information and bear many applications; for example, in the identification of functionally important residues within a sequence, such as residues that participates in enzymatic catalysis, ligand-binding and protein-protein interactions [3]. The evolutionary conservation analyses are especially valuable when combined with structural information. For example, residues that are part of a functional site, as a ligand-binding site or a catalytic pocket, are often scattered on the primary sequence. Thus, projection of the evolutionary conservation analysis onto the 3D-structure, as applied by the ConSurf server, could reveal conserved patches that point to the location of functional sites [6]. This information is valuable, for instance, to the structural genomic project by inferring function form structures [7].

Evolutionary conservation analyses also aid genetic studies with the discrimination of damaging vs. neutral mutations, which typically involve the substitution of conserved vs. variable residues, respectively [8]. When structural information is available, calculation of the solvent-exposed surface area [9] for each mutated residue complements the conservation analysis. For instance, a conserved and buried residue is often important for maintaining the structural fold, while a conserved and exposed residue is often of functional importance [10]. In case the mutated residue is evolutionary variable and surface-exposed, the substitution is most likely to be neutral. Such analyses aid in the examination of many mutations found in patients and in the understanding of their molecular effect and clinical relevancy, including in my own study described in chapter 1 (Landau and Ben-Tal (2008) *BBA-reviews on cancer* 1785(1) 12-31), and in my other studies that are not a part of this dissertation [11-19].

It is noteworthy that the evolutionary conservation profile is directly dependent on the set of homologous sequences used for the calculations [3]. For example, in my investigation of the tyrosine kinase receptors of the ErbB family, described in chapter 1, I compared between evolutionary conservation analyses that were based on a set of members of the ErbB family vs. a larger set of various kinases. This comparison highlighted specific residues and regions that are unique to the ErbB family among other kinases. Accordingly, by using different sets of homologous sequences, it is possible to reveal specific traits. Explicitly, I examined specific patches on the surface of kinases as potential sites for protein-protein interactions. My evolutionary conservation analyses pointed out that these patches are unique to members of the ErbB family among other kinases. This result was confirmed by the examination of the

physicochemical properties of the patches in the ErbBs and in other kinases [20] (Landau et. al. (2004) *Structure* 12, 2265-2275; the manuscript is attached as Appendix II).

Protein 3D-structures played a significant role in my examination of function and regulation mechanisms of transmembrane proteins. Unfortunately, due to experimental difficulties, we are still far beyond an inclusive structural determination of the transmembrane domain of proteins, especially from eukaryotes. As a result, although they comprise 15-30% of the eukaryotic genome, only 1% of the structurally determined proteins are membranal [21]. Consequently, computational tools are most valuable resources to obtain structural insights for these proteins [21]. For instance, in my work on Na^+/H^+ exchangers (Landau et al. (2007) *JBC* 282(52) 37854-63), described in chapter 2, I have predicted the 3D-structure of the human Na^+/H^+ exchanger isoform 1 (NHE1) using homology modeling tools [22] based on a structure of a functional homolog from *Escherichia coli* [23]. In addition, evolutionary conservation analyses had also been shown valuable to predict protein structures [21, 24-26]. Generally, residues that are important for maintaining protein structures are expected to be highly conserved within protein families sharing a similar fold. Specifically in helical transmembrane protein, as NHE1, the strategic and conserved locations within the structure are the interfaces between the transmembrane helices, in contrast to the evolutionary variable residues facing the membrane lipids or the ones located on the extra-membranal loops [21, 27, 28]. Accordingly, evolutionary conservation analyses have also been exploited to validate model structures [29]. Indeed, such analyses have played a crucial role in the validation of my proposed structure of NHE1.

Overall, my dissertation demonstrates the significant contribution of applying computational tools for the comprehensive understanding of protein mechanisms and particularly to guide hypothesis-driven experiments. Most prominently, I found that the integration of data from various computational approaches, along with gathering the relevant experimental information, is extremely valuable to raise models of regulations for protein families.

Chapter 1: Dynamic Equilibrium between Multiple Active and Inactive Conformations Explains Regulation and Oncogenic Mutations in ErbB Receptors

This chapter is based on two published manuscripts (Landau et. al. (2004) *Structure* 12, 2265-2275; Landau and Ben-Tal (2008) *BBA-reviews on cancer* 1785(1) 12-31)

Abstract

The ErbB growth factor receptor family members are key players in vital physiological and pathological processes. Like other receptor tyrosine kinases, the ErbBs are type I transmembrane proteins. In recent years the crystal structures of the extracellular and intracellular domains of some ErbBs have been determined. I integrated the available structural information with phylogenetic, biochemical, biophysical, genetic, and computational data into a suggested model for the regulation and activation of these receptors. According to the model, regulation is maintained by a dynamic equilibrium between monomeric and dimeric states in various conformations. Along this dynamic equilibrium, variations in the points of interactions within the dimers alter the activation state and ligand-binding affinities**.** The active state was recently shown to be associated with an asymmetric dimer of the kinase domains. That finding enabled me to elucidate, in molecular terms, the directionality observed in the activation process of ErbB heterodimers; it can explain, for example, the preferential activation of ErbB2 by ErbB1 over activation of ErbB1 by ErbB2. Sequence alterations that reverse this directionality lead to aberrant signaling and cancer. My model also offers molecular interpretations of the effects of various oncogenic alterations that interfere with the regulatory mechanism.

1. Introduction

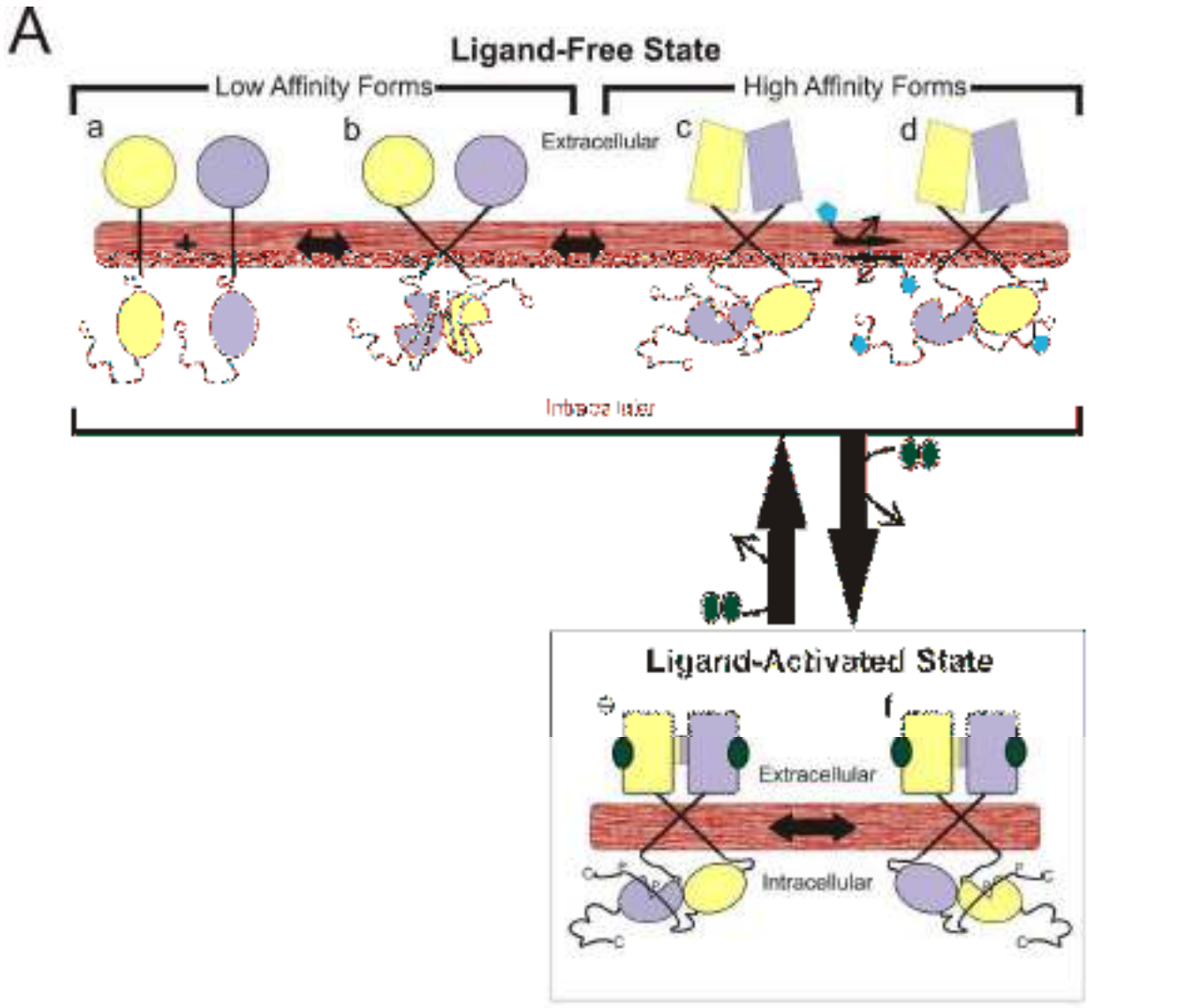
The four ErbB growth factor receptors are members of one of the most notorious protein families in cancer research, also known as the HER family. ErbB1, the first member to be discovered, is also known as the epidermal growth factor receptor (EGFR). The ErbB family is ubiquitously distributed throughout the animal kingdom [30], and plays an essential role in vital cellular processes such as proliferation, differentiation, migration, and apoptosis [31, 32] as well as in various pathologies [33].

The ErbBs belong to the receptor tyrosine kinase (RTK) super-family [34], all of which are type I transmembrane (TM) proteins that are activated by binding of extracellular ligands, such as the epidermal growth factor (EGF). Structurally they consist of an N-terminal extracellular ligand-binding domain, a single TM span, and a large intracellular domain that includes the catalytic kinase domain and a regulatory C-terminal domain (Fig. 1). Activation of the ErbBs induces phosphorylation of tyrosine residues that are located in the receptor's intracellular domain and serve as docking sites for other proteins, thereby allowing transfer of the signal into the cell. The complex signaling network of the RTKs in general, and of the EGFR family in particular, has been the focus of intensive research (see e.g., [32, 35]).

Interestingly, the second family member, ErbB2, lacks the capacity to bind ligands, and the third, ErbB3, has an inactive kinase [32]. Both are nevertheless vital to embryonic development, as are the EGFR and the fourth family member, ErbB4. This apparent paradox is explained by the fact that the basic functional unit in the signaling of the ErbB family, as in other RTKs, is a dimer. Thus, ErbB2 and ErbB3 operate by combining with other ErbBs to form heterodimers [32]. Phosphorylation of ErbB3 by its partner leads to a specific and unique signaling pathway, explaining the significance of ErbB3 despite its being catalytically defective.

The ErbBs mediate an essential cellular signaling network; consequently, their activation is subject to several layers of regulatory control [2]. In this work, I focused on the molecular details of the regulatory mechanisms that operate within the ErbB proteins. The regulatory role played by the extracellular and TM domains of the ErbBs has been extensively discussed. In contrast, most aspects of the regulatory mechanisms located in the intracellular domain have only recently been elucidated. Here I discuss the new developments and suggest a model of EGFR regulation, presented in Fig. 1, which integrates the multiple layers of control imposed by the various structural elements of this protein. I propose a mechanism of dynamic equilibrium, which is manifested by interconversions between different

conformations of the receptor in the ligand-free state. Binding of ligand disrupts this equilibrium and invokes a consecutive process that triggers activation. In the second section of this chapter, I explain the regulation imposed within and between the receptor's different domains, and this leads me to formulate the thermodynamic model presented in Fig. 1 and summarized in section 3. Differences and similarities in regulatory mechanisms among the various ErbBs are elaborated in section 4. In the fifth section, by referring to my regulatory model and an evolutionary conservation analysis of ErbBs from various species, I predict the molecular effect of cancer-causing mutations in the ErbBs. I note that the mechanisms of regulation of the EGFR might depend on the cell type and the stage of the cell cycle [36].


A
Ligand-Free State
Low Affinity Forms
High Affinity Forms
a
b
Extracellular
c
d
Intracellular
Ligand-Activated State
e
f
Extracellular
Intracellular

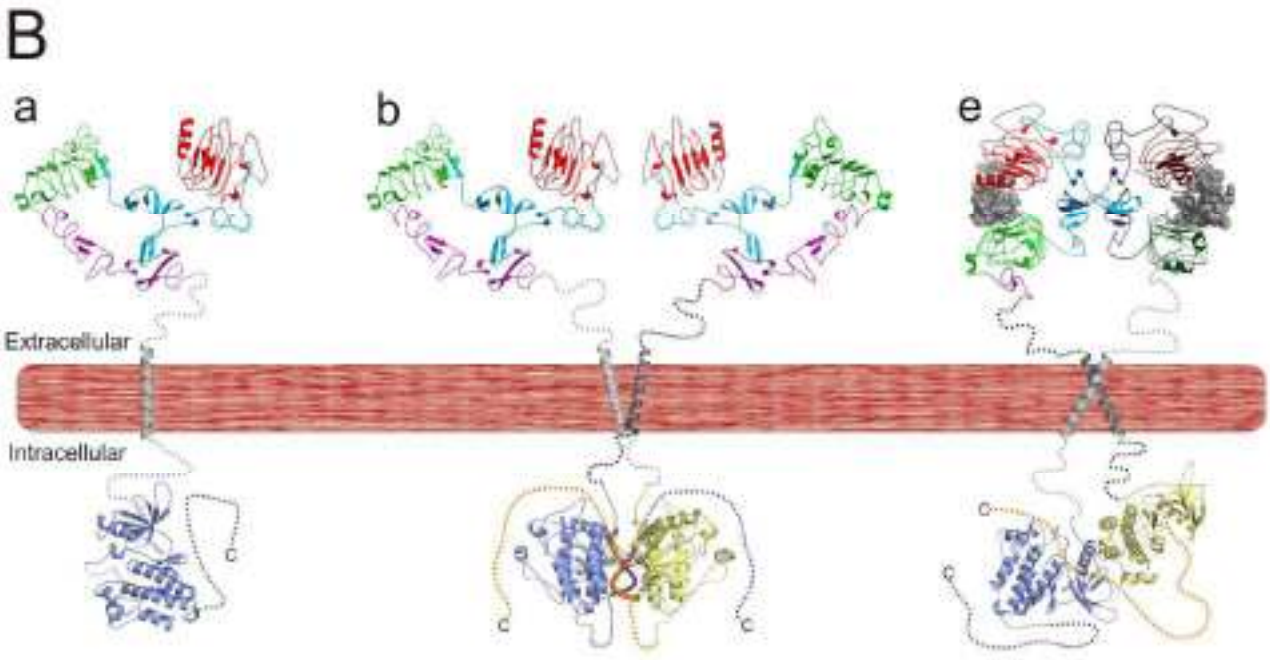

B
a
b
e
Extracellular
Intracellular

Fig. 1: Regulation of the EGFR

The EGFR is represented by its main structural elements, namely the extracellular part, the TM domain, and the intracellular part comprising the kinase and C-terminal domains. Location of the membrane is marked by the brown bar. **Panel A** presents a scheme of the regulatory mechanism maintained by interconversion between multiple forms. I note that all forms in the ligand-free state are in dynamic equilibrium with one another. Two monomers of the EGFR are displayed, colored yellow and purple. **Panel B** displays a ribbon representation of crystal structures available for specific domains of the EGFR. The fragmentary structures are schematically combined here to illustrate the entire receptor in a few conformations that correspond to panel A. Regions for which a crystal structure is not available are depicted by dashed curves. The crystal structures of the extracellular domain are colored according to its subdomains (I in red, II in cyan, III in green and IV in magenta; for the dimer in form 'e', one of the monomers is in darker shades for clarity). The EGF ligand is represented by gray space-filled atoms. The TM helices are colored in two shades of gray. The kinase domains are colored in yellow and purple as in panel A. A fragment from the C-terminal domain that is available in the crystal structure is displayed as a tube, colored in darker shades of the color of the kinase domain on the same monomer.

The ligand-free state is characterized by interconversion between many different conformations, while the scheme in panel A depicts only representative forms. In both panels, form 'a' stands for a monomeric conformation in which the kinase domain is inactive (residues 679−964; PDB entry – 2gs7; [37]). Form 'b' illustrates a dimeric form demonstrating the formation of an inactive intracellular complex (residues 672−995; PDB entry – 1m17; [38]), and an inactive TM complex (the coordinates were taken from [39]). The extracellular domains in forms 'a−b' are predicted to alternate between tethered and extended conformations that are embodied by an abstract conformation in panel A. The crystal structure of the extracellular domain (forms 'a−b' in panel B) depicts only the tethered conformation in the ligand-free state (residues 2−614; PDB entry – 1YY9; [40]). Forms 'a−b', in which the extracellular domains are separated, are predicted to have a relatively low ligand-binding affinity.

Based on experimental data (see main text) I predicted that an intracellular asymmetric dimer is one component of the high-affinity state. Forms 'c−d' display a dimeric conformation in which contact formation is mediated by the TM, the extracellular, and the intracellular domains. In these forms the extracellular domains assume an extended conformation allowing contact, which is likely to significantly increase the binding affinity of the receptors for the ligand. No structural information is available for such conformations of the extracellular domains, which are probably transient. The intracellular asymmetric dimer is comprised of activating (yellow) and activated (purple) monomers, which can potentially switch orientations (as illustrated in forms 'e−f'). The equilibrium between forms 'a', 'b' and 'c' is dependent on their inherent stability, while conversion to form 'd' is dependent on an unknown external factor that is depicted by a cyan polygon. I emphasize that the conformation assumed by form 'd' is entirely hypothetical.

Binding of the growth-factor ligands, depicted by green ovals in panel A (or gray spheres in panel B), induces conformational changes in the extracellular domains, including exposure of the 'dimerization arm' illustrated in forms 'e−f'. The ligand-bound extracellular domains (residues 2−512; PDB entry - 1IVO; [41]) are in extended conformation, and allow contact formation between the two monomers. In this ligand-activated state the TM complex assumes the active conformation (the coordinates were taken from [39]). The kinase domains

within the intracellular asymmetric dimer (residues 669−967; PDB entry – 2gs6; [37]) can potentially switch orientations, such that the activating and activated monomers exchange roles (forms 'e−f'; panel B illustrates only the structure corresponding to form 'e').

2. Mechanism of Regulation in the ErbB Family

The general scheme of the activation mechanism in RTKs involves binding of ligand to the extracellular domain, which leads to dimerization of the receptor followed by structural changes and phosphorylation of tyrosine residues in its intracellular domain [34]. Accumulating new evidence has resulted in fundamental extensions and modifications to the above basic mechanism in the ErbBs. These modifications are discussed here.

2.1. Regulation by Dimerization

The ErbBs were shown to be clustered in the membrane, and random collisions that could lead to the formation of dimers and other oligomers are therefore very likely to occur [36, 42, 43]. As a result, it is anticipated that an additional layer of regulation is needed in order to prevent activation induced by random dimerization. For example, the formation of stable inactive dimers or oligomers could serve as a mechanism for the required control. Studies have indeed shown that although dimerization (which requires spatial proximity between two ErbBs) is a necessary step toward ligand-induced activation, such proximity is not in itself sufficient to promote activation [36, 44-49].

Because the RTKs are multi-domain proteins, the term 'dimerization' should be used with caution, i.e., with implicit relevance to the location of interaction. As an example, mediation of contacts by the extracellular domain does not necessarily mean that the intracellular kinase domains directly interact, and vice versa. The location of contact formation between the two subunits has crucial implications for the activation state of the receptor (see Fig. 1).

As mentioned above, numerous studies have shown that the EGFR can form ligand-independent dimers, or 'pre-dimers' [44-51]. These dimers are probably inactive [36, 44, 50]. However, the nature of the specific interface within the dimers detected in those studies has yet to be fully elucidated. There are indications that all ErbBs display ligand-independent contact formation between their TM domains [47]. In addition, there is evidence that the cytoplasmic domain [44], and even more specifically, the kinase domain [45], are necessary for the formation of at least some of these ligand-independent dimers.

Binding of EGF to the pre-dimers on the cell surface displays much higher affinity than to the monomers [51] (see section 2.5). It was suggested that pre-dimer formation helps facilitate the formation of active dimers, which is not limited by diffusion along the plasma

membrane, even at low EGF concentrations [51]. Moreover, clustering of ErbBs in the membrane probably contributes to a rapid spread of the signal through receptors that are not bound to the ligand, but are activated by their adjacent active dimers [36].

2.2. Regulation by the Intracellular Domain

The kinase domain, which is located in the intracellular domain, is the catalytic unit in all RTKs. It is comprised of two subdomains, known as the N-terminal (N-) and the C-terminal (C-) lobes (Fig. 2). The phosphate donor (ATP), as well as the substrate designated for phosphorylation, bind in a cleft flanked by the two lobes. The ATP is held in position by a phosphate-binding loop (P-loop), which is located at the N-lobe. The substrate is positioned near the ATP's phosphates by the 'activation loop' located in the C-lobe. In most RTKs the activation loop assumes a closed conformation in the inactive state, thereby preventing substrate binding. Activation involves conformational changes and phosphorylation of the activation loop [2, 52, 53]. These events are coupled to a concurrent movement of another regulatory element, the αC-helix located in the N-lobe, into a final conformation that is catalytically competent. The actual catalytic process is performed by the highly conserved catalytic loop in the base of the active site located in the C-lobe. Catalysis requires that all of the above regulatory elements are positioned in the proper spatial arrangement.

Until recently, it was thought that the constant state of the kinase domain of the EGFR and other ErbBs is the active conformation. This assumption was based on structural data, which showed that although the activation loop is not phosphorylated, all of the regulatory elements of the EGFR's kinase domain are ready for catalysis [38]. That structure concurred with previous biochemical data showing that activation of the ErbB family is independent of their phosphorylated state [54]. Subsequently, when the crystal structure of the EGFR in complex with the large inhibitor GW572016 was determined, a putative inactive conformation of the kinase domain was observed. It was suspected, however, that this conformation was induced by the bulky inhibitor [20, 37, 55].

In contrast to the supposedly constant active conformation, biochemical data showed that the EGFR is not constantly active in cell membranes [34]; indeed, aberrant conditions (such as mutations in the EGFR) that lead to ligand-independent activation are known to be related to cancer. Thus, the phosphorylation-independent activation of the EGFR [54], as well as the lack of conformational regulation observed in the first crystal structure [38], prompted investigators to seek regulatory elements that govern the inhibition of the EGFR within the

intracellular domain [20, 37, 56], including my own study [20] (the manuscript is attached as Appendix II).

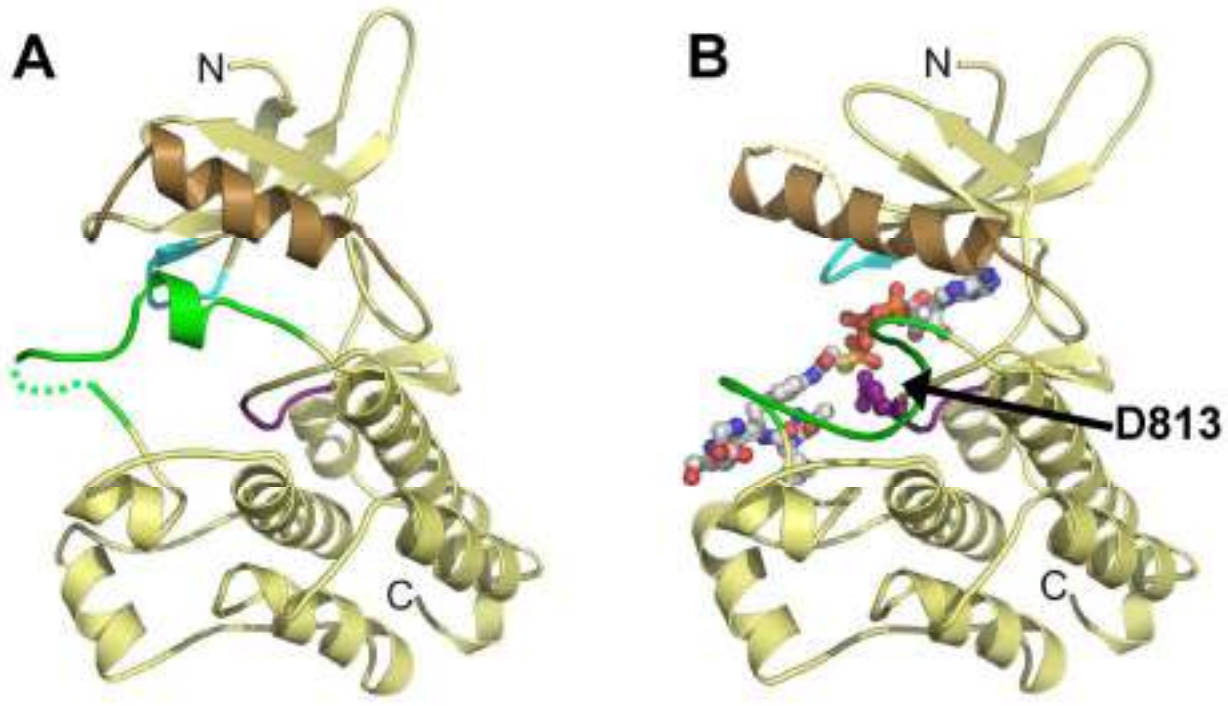


Fig. 2: Regulatory Elements in the Kinase Domain

The kinase domain of the EGFR (residues 685−957) is depicted in a yellow ribbon representation, in its inactive **(A)** (PDB entry – 1gs7; [37]) and active **(B)** (PDB entry – 2gs6 ; [37]) conformations. Regions that are not ordered in the crystal structures are depicted by dashed lines. Regulatory elements are depicted by different colors: the activation loop (residues 831−852) in green, the αC-helix (residues 729−745) in brown, the phosphate-binding loop (residues 695−700) in cyan, and the catalytic loop (residues 812−818) in purple. In panel B, the catalytic residue Asp813 (in purple) and the ATP analog-peptide conjugate (colored by atom type: carbon in white, nitrogen in blue, oxygen in red, phosphate in orange, and sulphur in yellow) are displayed in a balls-and-sticks representation.

2.2.1. The Inactive State

In a recent study by John Kuriyan and his co-workers it was shown that the kinase domain of the EGFR indeed also possesses an inherently inactive conformation [37]. Those authors suggested that the constitutive active conformation of the EGFR that was seen in the first crystal structure [38] was induced by intermolecular contacts within the crystal. A mutation in the kinase domain that disrupted such contacts led to the identification and determination

of an inactive kinase conformation [37], which was very similar to the GW572016-bound structure [55].

The inactive conformation of the EGFR resembles those of the cytoplasmic tyrosine kinase Src and the serine/threonine cyclin-dependent kinases (CDKs) [37]. In this state the αC-helix, together with the loop preceding it, forms contacts with the N-terminus of the activation loop, combining this threesome in an inactive conformation (Fig. 2A). Consequent displacement of the αC-helix and the activation loop prevents the proper positioning of catalytically important elements (Fig. 2B). Any interference with the packing of the activation loop against the αC-helix and its preceding loop, caused for example by mutations, would lead to increased auto-kinase activity and cell transformation [57]. These mutations will be discussed in section 5.

The C-terminal domain serves as an auto-inhibitor via direct contacts with the kinase domain within a dimer

By and large, ErbB signaling and part of its regulation are mediated via the C-terminal domain. Following activation, phosphorylated tyrosines on the C-terminal domain serve as docking sites for the subsequent proteins in the signaling cascade, as well as for regulatory proteins that control processes such as internalization and degradation [32, 33, 58]. I recently suggested that a fragment from the C-terminal domain of the EGFR, immediately following the kinase domain, serves as an inherent auto-inhibitor of the receptor via its direct contacts with the kinase domain [20].

The structures of the intracellular domain of the EGFR contain the kinase domain (residues 672−957; amino-acid numbers throughout the text are based on mature EGFR) and a fragment from the C-terminal domain (residues 958−995) [37, 38, 55]. According to my earlier analyses, the kinase domain of the EGFR displays two surface patches of positive electrostatic potential, one in the N-lobe and the other in the C-lobe, both located on the back side of the active site [20]. On the other hand, the C-terminal fragment that is present in the structure (residues 979−995) is highly negatively charged.

The structures of the wild-type (wt) EGFR kinase domain reveals a crystallographic symmetric dimer in which contacts between the two kinase domains are mediated by the two C-terminal fragments (Fig. 3) [37, 38]. Within this dimer, each kinase domain contacts the two C-terminal fragments via its two positively charged surface patches. In addition to this charge complementarity, my previous analyses showed that the water-accessible surface areas between each kinase domain and the two C-terminal fragments in this complex are

extremely large (1419Å and 1048Å for the N- and C-lobe interfaces, respectively) [20]. Moreover, although the two kinase domains in the complex barely interact directly, they are connected via the C-terminal fragments through a network of salt bridges and hydrogen bonds [20]. Overall, this complex appears to be extremely stable and therefore has the potential to prevent ligand-independent activation induced by random contacts. I also observed that one of the phosphorylation sites on the C-terminal domain (Tyr992) is buried within the complex and is inaccessible to phosphorylation [20]. Thus, I speculated that this complex represents an inactive form of the receptor (form 'b' in Fig. 1).

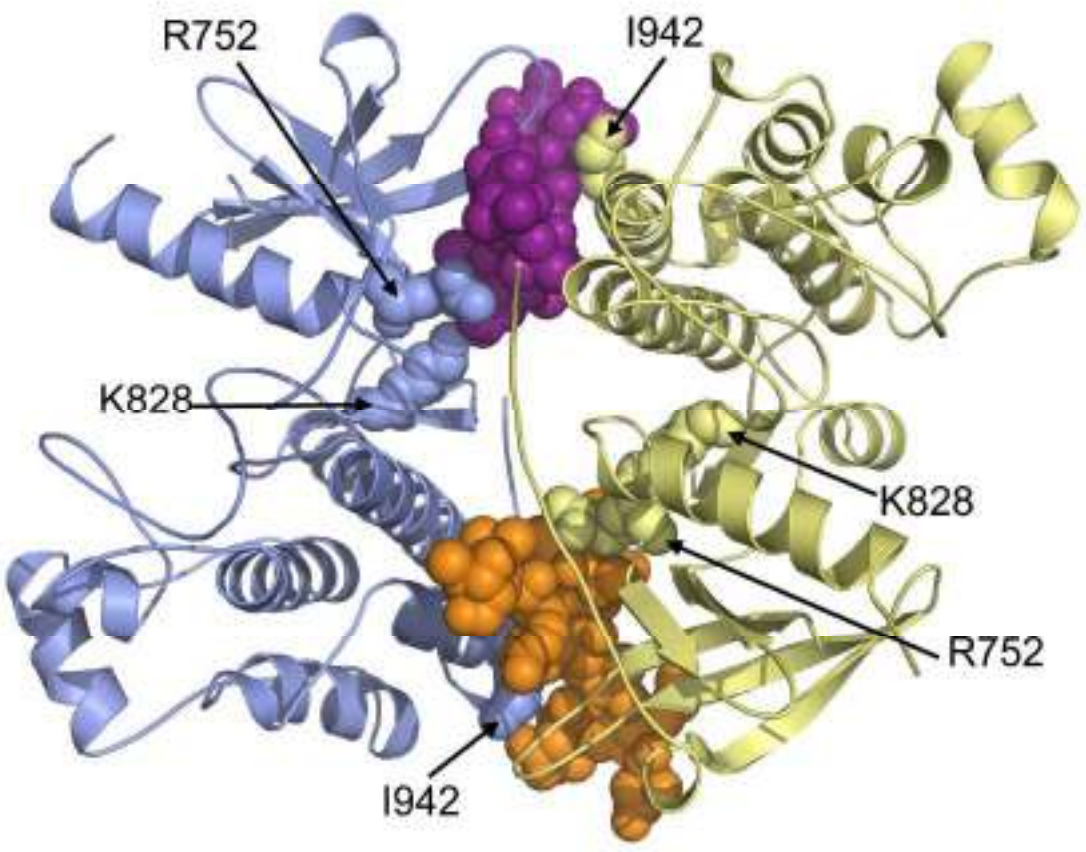


Fig. 3: The Symmetric Inactive Dimer of the Kinase Domain

Within the inactive symmetric crystallographic dimer, contacts between the two kinase domains (residues 672−966; depicted in yellow and light purple ribbon representations) are mediated by their following two C-terminal fragments (residues 982−964, depicted as space-filled atoms and colored in darker shades of the color of the kinase domain on the same monomer). Residues on the kinase domain that are in contact with the C-terminal fragment, and whose replacement was implicated in heightened basal activity (Lys828 and Ile942) or was found in cancer cells (Arg752), are depicted as space-filled atoms. Coordinates were taken from PDB entry – 1gs2 [37].

In accordance with the above model supporting the role of the C-terminal domain as an auto-inhibitor of the EGFR, variations in the C-terminal domain in naturally occurring retroviral oncogene variants (v-ErbB) are associated with malignant diseases because of their increased rate of auto-phosphorylation [59-65]. More specifically, engineered deletions or mutations in the C-terminal fragment in contact with the kinase domain in the crystal structure display heightened catalytic activity [37, 58]. Such variations are predicted to cause destabilization of the inhibitory interactions between the kinase and the C-terminal domains, promoting activation [20]. Furthermore, specific replacements of four negatively charged residues on the C-terminal fragment (residues 979−982) by their polar equivalents resulted in higher auto-kinase activity and partial transforming phenotype [58], supporting the role of charge complementarity in the inhibition by the C-terminal domain. Correspondingly, substitution of alanine for the positively charged Lys828 (K828A), which participates in intramolecular interactions with the C-terminal domain (Fig. 3), resulted in higher basal activity of the receptor [37]. Interestingly, this residue participates in the above mentioned polar network of interactions that mediates the contacts within the symmetric dimer [20]. These results indicate that the C-terminal domain indeed participate in inhibition via direct interactions with the kinase domain.

It was previously suggested that the monomeric inactive state of the kinase domain is stabilized by direct contacts between the kinase and the C-terminal domains [55]. However, *in-vitro* experiments, conducted using a truncated form of the intracellular domain of the EGFR (residues 672−998) at a low concentration that mimics the monomeric form of the receptor, showed that the catalytic activity of the kinase domain is not altered by further truncation of the C-terminal fragment (residues 965−998) [37]. This implies that the C-terminal fragment does not participate in stabilizing the monomeric inactive conformation. On the other hand, at a high concentration of the EGFR that corresponds to the concentration at which dimers are formed, truncation of the C-terminal fragments greatly increased the inherent activity of the kinase domain [37]. Evidently, therefore, the effect of the C-terminal fragment is manifested only when the EGFR forms dimers, corresponding to my earlier suggestion [20]. Interestingly, a mutation (I942E) in a residue which, according to the symmetric crystal dimer, contacts the C-terminal fragment of the adjacent monomer (Fig. 3), led to higher basal activity of the receptor [37], supporting the biological relevance of this dimer and its role in auto-inhibition.

Overall, the inactive state of the EGFR probably involves an intracellular dimer that might resemble the symmetric dimer seen in crystals, as I suggested. This crystallographic complex,

however, evidently does not play a role in the active state, as mutations in residues located on its interfaces did not alter the ligand-induced kinase activity of the EGFR [37].

2.2.2 The Active State

The exact conformation of the active state of the intracellular domain, after eluding investigators for years, was finally revealed by John Kuriyan and his co-workers [37]. The crystal lattice of the EGFR includes, in addition to the symmetric dimer discussed in section 2.2.1, an asymmetric dimer [37, 38]. Through a series of comprehensive and elegant experiments, Kuriyan and his co-workers showed that this asymmetric dimer represents the active form of the intracellular domain, in which one kinase activates its partner [37] (forms 'e−f' in Fig. 1).

This asymmetric dimer closely resembles the complex between cyclinA and activated CDK2 [37] in which cyclinA serves as an activator of CDK2 [66]. In the asymmetric dimer of the EGFR, one kinase maintains contact with its partner through the C-lobe in a way that mimics the role of cyclinA. Thus, by analogy, I refer to it as the 'activating' monomer (the yellow kinase in form 'e' in Fig. 1). The second kinase, which forms the dimeric interface via its N-lobe, mimics CDK2, and I therefore refer to it as the 'activated' monomer (the purple kinase in form 'e' in Fig. 1). Interaction via the N-lobe leads to displacement of all the regulatory elements into the catalytically active conformation.

2.3. Regulation by the Cytoplasmic Juxtamembrane Domain

The region immediately following the TM domains, termed the juxtamembrane (JM) domain (residues 645−672), might also play a role in regulation [56, 67, 68]. An EGFR mutant that sustained a deletion of the JM segment (residues 645−657) displayed similar basal activity to that of the wt EGFR, but lost its ability to interact with a second EGFR molecule and to undergo ligand-induced phosphorylation [67]. Accordingly, this region can be expected to play a role in the active dimeric state, but not in auto-inhibition.

The JM segment carries a positive net charge and was therefore suggested to bind to the negatively charged inner leaflet of the membrane [56]. It was further suggested that binding of the negatively charged Ca^{2+}/calmodulin complex would repel the JM domain from the membrane, with subsequent possible rearrangement of the kinase domain into an active state [56, 68]. Initial activation of the EGFR leads to a transient increase in the cellular level of free Ca^{2+}, thereby activating the $Ca^{2+/}$calmodulin complex [56, 68]. Taken together, therefore, these studies suggested that binding of the Ca^{2+}/calmodulin complex to the JM domain might

serve to amplify the initial signal by subsequently activating large numbers of additional receptors.

2.4. Regulation by the TM Domain

The TM domain of the ErbBs is more than a passive peptide anchoring the receptors to the membrane; it also serves as an additional layer of regulation [29]. The ErbB TM domain has an inherent tendency to associate in the membrane in two different forms that are mediated by two different dimerization motifs [39, 47]. Based on energy considerations, earlier computational analysis assigned these dimerization motifs in ErbB2 to the inactive (form 'b' in Fig. 1) and active (forms 'c−f' in Fig. 1) states and suggested that the TM dimers could switch between these two states by rotating through 120° in a screw-like motion [39, 69]. This and other studies suggested a rotation-coupled activation mechanism in which ligand binding induces a rotational change in the TM helices from the inactive to the active conformation [39, 70]. The change in conformation is translated into reorientation of the intracellular kinase domains, leading to activation (Fig. 1) [20, 39, 48, 49, 70]. This molecular-switch model provided an explanation, at the molecular level, for known ErbB2 mutants or naturally occurring variants [39]. For example, the oncogenic mutation in the rat ErbB2 (*Neu*), which is located within one of the dimerization motifs, was predicted to induce a shift toward the active conformation and hence towards enhanced activation [39]. The observed existence of oncogenic mutations in the TM domain further supports a regulatory role for this domain [39].

It is interesting to note that in the predicted inactive conformation, the TM-segments interact through the dimerization motif located at the C-terminal part of the helix, resulting in close proximity of its intracellular ends (~9Å), as illustrated in form 'b' in Fig. 1. Rotation to the active state, mediated by the N-terminal dimerization motif (forms 'c−f' in Fig. 1), would impose a larger distance (~19Å) between the cytoplasmic ends of these helices [39]. Correspondingly, the N-termini of the putatively inactive intracellular dimer are much closer to each other (~20Å; form 'b' in Fig. 1B) than in the active asymmetric dimer (~50Å; form 'e' in Fig. 1B). In addition, switching of the TM domains to the active state brings their extracellular ends closer to one another, which corresponds to oncoming of the extracellular domains, and brings them closer to contact formation (forms 'c−f' in Fig. 1). Thus, in accordance with the rotation-coupled activation mechanism, the distance between the cytoplasmic and extracellular ends of the TM helices allows leverage that controls the positions of the intracellular and extracellular domains, respectively.

2.5. Regulation by the Extracellular Domain

In recent years, a growing body of structural information has led to substantial progress in understanding the mechanisms of ErbB regulation by their extracellular domains. The ligand-free structures of the extracellular domains of EGFR [40], ErbB3 [71], and ErbB4 [72] display a tethered auto-inhibited state, while the EGFR's ligand-bound extracellular domain [41, 73] shows an extended conformation that allows contact formation between two subunits (Fig. 1B). Interestingly, the 'orphan' (ligand-less) receptor ErbB2 resembles the ligand-bound active conformation of the EGFR [74, 75].

The role of the extracellular domain in regulation of the ErbBs, excluding ErbB2, is to impose a ligand-mediated activation [36, 76]. Correspondingly, truncation of the extracellular domain, as seen in viral ErbB variants that are related to oncogenic transformation, leads to ligand-independent activation of the receptors [61]. These findings also suggest that other domains, namely the TM and intracellular parts, possess an inherent ability to interact and to become activated.

The structure of the extracellular domain has been described in detail in a few comprehensive reviews (e.g., [76, 77]). In brief, it comprises four subdomains, termed I−IV (Fig. 1B). In the inactive conformation, subdomains II and IV interact to lock the structure of the extracellular domains of ErbB-1, -3, and -4 in a tethered, putatively inactive conformation [78] (forms 'a−b' in Fig 1B). Biochemical experiments indicate, however, that other elements in the extracellular domain are probably also involved in the auto-inhibition [35, 79]; thus, the available structures do not yet provide the whole story of inactivation [36]. In the extended dimeric ligand-bound conformation, each growth-factor-ligand binds one extracellular domain through subdomains I and III [41, 73] (form 'e' in Fig. 1B). Contacts between the two EGFR extracellular domains are mediated solely by the receptors, mainly by a loop (the 'dimerization arm') in subdomain II [41]. Mutations in this loop prevent ligand-induced activation [41, 73].

Measurements of the binding affinity of EGF to its receptor pointed to the existence of two different EGFR populations on the cell surface, each with its own ligand-binding affinity [36, 76]. It was proposed that 92−95% of the receptors exhibit low-affinity binding, with a Kd of approximately 6−12 nM. These receptors were predicted to assume the tethered conformation of the extracellular domain, in which the two subdomains responsible for making contact with the ligand are far apart, allowing the ligand to make contact with only one of two interfaces [71, 76, 78]. Some 2−5% of the receptors were proposed to bind EGF with high affinity, with a Kd below 0.1 nM, and these were considered to assume the extended

conformation that brings the two ligand-binding subdomains closer [51, 71, 76, 78]. However, mathematical models describing the kinetics of interactions between EGF and its receptor showed that there is no direct correlation between the conformations seen in the crystal structures and the two different binding-affinity populations [80]. Recent findings suggested that the low-affinity population actually corresponds to an ensemble of conformations that are related to interconversion between the tethered and the extended forms of the extracellular domain [35]. Conformation of the high-affinity state is less well characterized, and are thought to involve additional elements, such as other proteins [81, 82] (form 'd' in Fig. 1A) as well as cell-regulatory mechanisms such as endocytosis and degradation [80]. Moreover, the high-affinity binding is totally dependent on specific regions within the intracellular domains [35, 82, 83] (see also section 2.6.).

Overall, it seems that the low- and high-affinity binding modes, which appear macroscopically as two distinct populations that do not interact, rather represent interconversion between multiple conformations [82]. Some of these conformations are presented in the dynamic equilibrium presented in Fig. 1, where each form displays its own distinctive binding affinity and specific probability, based on its inherent stability. Conformations that display lower affinity for the ligand are probably more stable energetically than the higher-affinity conformations; this would explain the above macroscopic estimation that the binding affinity of 92−95% of the receptors is low. Nevertheless, the ligand preferably traps the less stable, high-affinity conformations. In the high-affinity state, the specific orientation of the TM and intracellular domains probably triggers contact formation between the monomers' extracellular domains (forms 'c−d' in Fig. 1A), although not necessarily in exactly the same manner as in the ligand-bound dimer (forms 'e−f' in Fig. 1A) [84]. Accordingly, a recent single-molecule analysis of EGF binding on the surface of living cells predicted that the macroscopic high-affinity population is represented by pre-dimers, which bind EGF at a rate two orders of magnitude higher than that of binding by the monomeric receptor [51].

As a result of ligand binding the formation of contacts between the two monomers is further stabilized, thereby shifting the equilibrium toward the ligand-activated state [51, 76] (forms 'e−f' in Fig. 1A). Recent findings predicted the formation of a kinetic intermediate following binding of the first EGF and prior to binding of the second [51]. Thus, binding of one EGF to the pre-dimer probably induces allosteric conformational changes that enhance the binding of the second EGF, hence displaying positive cooperativity [51].

Somatic or engineered alterations in the extracellular domain can potentially shift the equilibrium (depicted in Fig. 1) by thermodynamically stabilizing or destabilizing specific forms, thereby modifying the macroscopically measured ligand-binding affinity. For example, interference with the tethered conformation by mutations would lead to a higher average of the low-affinity binding mode, probably by shifting the equilibrium toward the extended conformations [71, 78]. On the other hand, alterations in the dimerization interface in the extracellular domain could destabilize contact formation and weaken the interaction between the EGFR and EGF, eventually lowering the inherent binding affinity [79].

2.6. Interactions within the Intracellular Domains Affect Ligand-Binding Affinity

As mentioned in section 2.5, the macroscopically measured low-affinity mode of binding of ligand to the EGFR is actually an average of the binding affinities of few different conformations that undergo interconversion. Interestingly, this average binding affinity could be modulated by intracellular elements, probably through shifts between conformations in the equilibrium scheme. For example, deletion of a fragment from the C-terminal domain (residues 984−996) resulted in an increase in the average binding affinity of the low-affinity mode by approximately threefold [83]. A similar effect was reported for a Y992E mutation within this region [82]. The same C-terminal fragment mediates the putatively inactive dimer (Fig. 3) discussed in section 2.2.1. I previously suggested that deletion of this fragment would destabilize the dimer [20] (form 'b' in Fig. 1), allowing more abundant ligand-independent formation of the asymmetric intracellular dimer (forms 'c−d' in Fig. 1A), which displays high-affinity binding. Overall, the effects of the mutations would result, on average, in higher binding affinity [82, 83].

The macroscopically measured high-affinity mode, representing only 2−5% of the receptors, probably reflects thermodynamically unstable high-affinity conformations that undergo interconversion with the more stable low-affinity forms. A clue to the conformations of the high-affinity state comes from engineered alterations to the receptors, where truncation of the entire intracellular domain resulted in the complete loss of high-affinity binding [35, 85]. More specifically, two regions were described as particularly crucial for high-affinity binding, namely the segment comprising residues 921 to 940, and the last 63 residues at the C-terminal end of the protein. Deletion of either of these regions completely abolishes the high-affinity binding mode [83 , 85]. It was suggested that the C-terminal end of the intracellular domain might comprise a binding site for external regulators that stabilize a high-affinity state [82] (form 'd' in Fig. 1A). Deletion of this region would prevent the

formation of this high-affinity form. Interestingly, I note that the other segment related to high-affinity binding (residues 921−940) comprises a large part of the interface within the active dimer. That finding led me to suspect that at least some of the high-affinity conformations could be correlated with the formation of the active asymmetric intracellular dimer, as represented by forms 'c−d' in Fig. 1A. I suggest that deletion of this region might disrupt the active intracellular dimer, thereby shifting the dynamic equilibrium toward lower-affinity forms. This model raises a question: does part of the high-affinity class (form 'c' in Fig. 1A) represent a ligand-independent active state that might be responsible for the basal activity of the receptor?

It should be emphasized that the high-affinity state is still only poorly understood and that not all of the experimental results obtained to date in this field can be satisfactorily explained. Accordingly, the high-affinity forms presented in Fig. 1 are provisional, and probably represent only a fraction of the conformations comprising the high-affinity state.

3. A Multilayered Model for the Regulation of EGFR Activity

Research on EGFR regulation recently reached a critical stage, at which the accumulated experimental data and models are converging in support of a unified mechanism of receptor activation. In this scheme, each of the domains in the EGFR constitutes an additional level in the regulatory mechanism. In the ligand-free state the receptors shift between multiple monomeric and pre-dimeric forms. In the monomeric form, the kinase domain assumes an inherently inactive conformation that prevents catalytic activity (form 'a' in Fig. 1). In the dimeric form, the TM helical segments preferentially interact through the dimerization motif located at the C-terminal part of the helix, resulting in close proximity of the intracellular ends to one another, but a substantial distance between the two extracellular ends (form 'b' in Fig. 1). In this conformation the two intracellular domains interact with one another, for example they form the inactive crystallographic dimer in which the N-termini are close to each other while the extracellular domains are separated. This dimeric conformation is catalytically inactive. In both the monomeric and the dimeric inactive states (forms 'a−b' in Fig. 1) the extracellular domain might assume either tethered or extended conformations that undergo interconversion, each displaying its distinctive affinity for the growth factor ligands.

A small energy barrier separates the stable inactive form from the less stable active form of the TM dimer, mediated by the C- and the N-terminal motifs, respectively [39]. Thus, I assume that, while in the ligand-free state, the active TM dimer could be populated to some

extent in which the cytoplasmic ends of the monomers are far apart and the extracellular ends are close together (forms 'c−d' in Fig. 1A). This would lead to destabilization of the inactive intracellular dimer, prompting the formation of the active asymmetric intracellular dimer in which the N-termini are distant from one another, and hence possibly to transient activation of the receptor (form 'c' in Fig. 1A). This state could also be stabilized by binding of regulatory proteins (form 'd' in Fig. 1A). At the same time, the close proximity of the extracellular ends of the TM segments would lead to contact formation between the extracellular domains. These conformations would typify the high-affinity ligand-binding mode.

Overall, the ligand-free state is characterized microscopically by dynamic equilibrium between different conformations (forms 'a−d' in Fig. 1). Because the low-affinity forms ('a−b') possess greater thermodynamic stability than the high-affinity forms ('c−d'), the equilibrium results in what is macroscopically measured as a large population of receptors displaying a low-affinity binding mode. On the other hand, receptors that assume the high-affinity forms ('c−d') are responsible for what appears macroscopically as a small population of high-affinity receptors.

Although contact formation between the extracellular domains in the ligand-free state (forms 'c−d' in Fig. 1A) induces high-affinity binding, these forms do not necessarily resemble the ligand-bound dimers (forms 'e−f' in Fig. 1A) [84]. Binding of the growth factor ligand stabilizes an active extracellular domain dimer (depicted by exposure of the dimerization arm in forms 'e−f' in Fig. 1). This in turn stabilizes the active conformation of the TM helical pair that promotes stable formation of the active intracellular dimer. Because this intracellular dimer is asymmetric, it could presumably be composed of two possible monomeric orientations (forms 'e−f' in Fig. 1). Overall, the equilibrium is strongly shifted toward the catalytically active state.

4. Are all ErbBs Regulated in the Same Way?

The ErbB family is ubiquitously distributed throughout the animal kingdom. There are four family members in vertebrates, whereas invertebrates have only one. The first gene duplication, in the early divergence of the vertebrates, generated the ErbB1/ErbB2 and ErbB3/ErbB4 precursors, each of which underwent a second gene duplication event to generate the four receptors present in vertebrates [30]. The unique sequence of each of these four isoforms dictates their particular regulation and signaling. Some examples are described

below. The presence of only one ErbB isoform in invertebrates implies that most aspects of its regulatory mechanisms differ from those in vertebrates [30]. Thus, I focus on the evolutionary conservation of ErbBs in species that diverged after the generation of four isoforms.

4.1. Regulation of the Ligand-less ErbB2 via the Extracellular Domain

ErbB2 is unique among the ErbBs in that it lacks the layer of inhibitory regulation, which is provided by the extracellular domain and imposes ligand-induced activation. Ligand-free ErbB2 assumes the extended, otherwise ligand-bound conformation of the EGFR [74, 75]. Accordingly, the extracellular domain of ErbB2 exhibits sequence variations in residues that participate in inter-subdomain interactions stabilizing the tethered conformation and which are conserved in ErbB-1, -3 and -4. Moreover, ErbB2 undergoes additional inter-subdomain interactions that mimic ligand-mediated contacts in the other ErbBs. Correspondingly, conserved residues that participate in ligand binding in other ErbBs are replaced in ErbB2 by residues that can participate in the inter-subdomain interactions [76]. Regardless of the constantly extended conformation of its extracellular domain, ErbB2 does not show a tendency to form homodimers. This might be due to electrostatic repulsion by negative charges on the predicted interface [76]. Nevertheless, ErbB2 efficiently forms heterodimers with other ErbBs [86]. Like ErbB2, ErbB3 is also reluctant to form homodimers, even in response to its ligand, neuregulin [87]. Nevertheless, neuregulin readily induces the formation of the most prevalent and highly efficient ErbB3−ErbB2 heterodimers [86].

4.2. Regulation of the Kinase-Dead ErbB3 via its TM Domain

As mentioned above, the TM segments of most ErbBs display two dimerization motifs and it was suggested that they correspond to the active and inactive conformations [39]. Interestingly, the kinase-dead ErbB3 contains only the dimerization motif that corresponds to the active state according to this suggestion [39]. Thus, ErbB heterodimers that contain ErbB3 have probably lost the layer of negative regulation imposed by the TM domains and hence more readily favor the active conformation. Together with the loss of regulation in the extracellular domain of ErbB2, this could explain the observation that ErbB2−ErbB3 heterodimers are extremely potent [86].

4.3. Regulation by the Intracellular Domain

4.3.1. Loss of Intramolecular Regulation in ErbB3

The monomeric inherent inactive conformation of the kinase domain is stabilized by packing of the activation loop against the αC-helix and its preceding loop [37], for example via contacts between residues Ile735 and Leu837 and between Met742 and Leu834 (Fig. 4B). It is interesting to note that these residues, which are putatively important for the packing, show some variation in ErbB3 (Fig. 4C), whereas they are conserved among other ErbBs. Mutations in these positions have also been found to correlate with malignancy, presumably because of destabilization of the active state; this is discussed further in section 5.1.1. Overall, my structural and sequence analyses suggest that there are fewer constraints on stabilization of the inherent inactive conformation of the kinase domain of ErbB3 than of other ErbBs. This observation is consistent with the reported lack of intramolecular activity in the kinase-dead ErbB3 [86]. Nevertheless, one have to take into account the activation of another ErbB member by heterodimerization with ErbB3. Accordingly, as described below, the regulation imposed by intermolecular interactions (e.g., via formation of an inactive intracellular dimer) is still maintained in ErbB3.

4.3.2. Negative Regulation of All ErbBs by an Inactive Intracellular Dimer

The symmetric and presumably inactive crystallographic EGFR dimer (Fig. 3) shows charge complementarity between the kinase and the C-terminal domains at the interfaces, as described in section 2.2.1. The positively charged residues on the interfaces in the kinase domain are highly conserved in all four ErbBs. Accordingly, my previous analyses, based on predicted structures, demonstrated that all of the human ErbBs display positively charged patches at the corresponding regions [20]. Evolutionary conservation analyses and electrostatic calculations show that those interfaces are not common to other tyrosine kinases [20]. In addition, the C-terminal fragment in contact with the kinase domain in the crystal structure in all ErbBs contains mostly negatively charged residues, even though the entire C-terminal domain is not generally conserved [30]. This implies that the electrostatic complementarity between the kinase domain and the C-terminal fragment might be important for controlling the activity of homo- and heterodimers in the entire ErbB family.

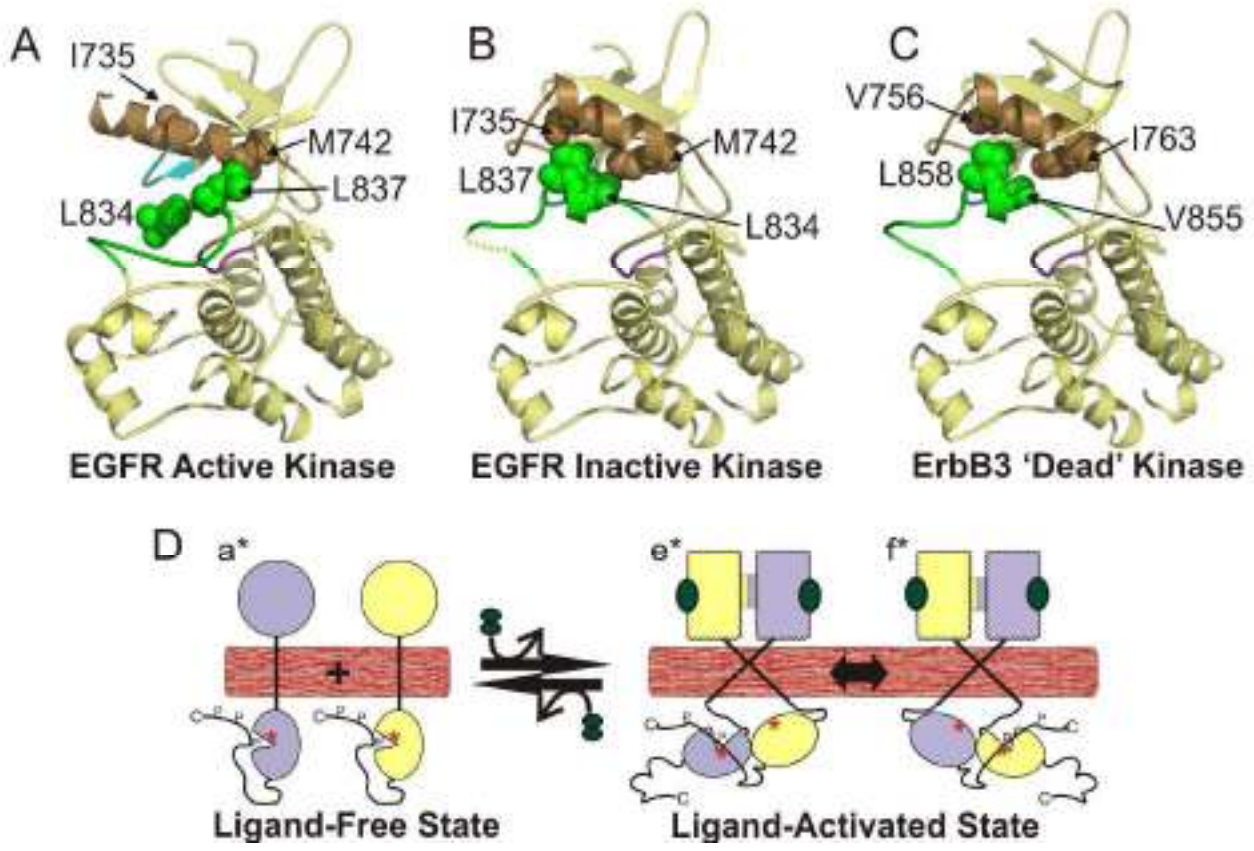


Fig. 4: Destabilizing Mutations in Regulatory Regions

The kinase domain of the EGFR (residues 685−957) is depicted in a yellow ribbon representation. Regulatory elements are depicted by different colors: the activation loop (residues 831−852) in green, the αC-helix (residues 729−745) in brown, the phosphate-binding loop (residues 695−700) in cyan, and the catalytic loop (residues 812−818) in purple. A disordered region within the activation loop is depicted by a dashed curve. Residues shown by mutagenesis to be crucial for stabilizing the inherent inactive conformation of the kinase domain of the EGFR are depicted as space-filled atoms. **(A)** Active kinase domain of the EGFR (PDB entry – 2gs6; [37]). **(B)** Inactive kinase domain of the EGFR (PDB entry – 1gs7; [37]). **(C)** Model of the catalytically inactive kinase domain of ErbB3. **(D)** Scheme illustrating alterations of the regulatory model (shown in Fig. 1A) by mutations, e.g., in residues Ile735, Met742, Leu834, and Leu837, which are shown in panels A−B. The mutations (indicated by red stars) cause destabilization of the inherent inactive conformation, leading to the formation of an active kinase domain whose phosphorylation is independent of the ligand or intermolecular contacts (form a*). Binding of ligand induces formation of the active dimer (forms e*−f*) that can potentially phosphorylate in *trans* and hence further increase the rate of auto-phosphorylation.

4.3.3. Functional and Structural Asymmetry of the Active State

The active state of the intracellular domain is characterized by formation of an asymmetric dimer of the kinase domains, in which one monomer activates the other [37]. It was previously shown that ErbB heterodimers could phosphorylate in *trans*, i.e., that the active

monomer could phosphorylate its neighbor within the dimer, as for example in the apparent phosphorylation of the kinase-dead ErbB3 by its partner [32]. Phosphorylation of both kinases would necessitate a switch between the orientations of the two subunits (illustrated by the transition between forms 'e' and 'f' in Fig. 1A). The structural asymmetry implies that at any given time only one of the kinases is catalytically active (represented by the purple kinase in form 'e' in Fig. 1) and can phosphorylate the C-terminal domain of its partner, namely the activating monomer (represented by the yellow kinase in form 'e' in Fig. 1). Interestingly, some ErbB heterodimers show preference for the formation of a specific orientation within the asymmetric dimer, as discussed further below.

The kinase-dead ErbB3 favors the role of activating monomer

In heterodimers that include ErbB3, only one direction of interaction, in which ErbB3 is the activating subunit, is functionally productive. Evolutionary conservation analysis of the ErbBs suggests that the interface of the activating monomer within the asymmetric dimer (yellow kinase in form 'e' in Fig. 1) is conserved among all four ErbBs. In contrast, most residues comprising the interface of the activated monomer (purple kinase in form 'e' in Fig. 1) are conserved only in ErbB-1, -2, and -4, while ErbB3 shows sequence variations. Presumably, therefore, an interface in which ErbB3 assumes the position of the activated subunit is energetically less stable.

ErbB2 favors the role of activating monomer within EGFR−ErbB2 heterodimers

Generally speaking**,** ErbB2 requires a catalytically active EGFR for its signaling, both in normal and in cancerous conditions [88, 89]. Within EGFR−ErbB2 dimers the EGFR is mostly found in the position of the active monomer (purple kinase in form 'e' in Fig. 1) that phosphorylates the C-terminal domain of ErbB2 (yellow kinase in form 'e' in Fig. 1), leading to cellular signaling that is specific to ErbB2. Accordingly, signal transduction mediated by ErbB2 is blocked by the EGFR-specific kinase inhibitor gefitinib [90, 91].

ErbB2, uniquely among ErbBs, possesses an additional layer of regulation, mediated by molecular chaperones (primarily Hsp90) that normally help to stabilize the receptor on the cell surface [92]. Hsp90 binds to ErbB2 and restrains ErbB-signaling by limiting its formation of heterodimers with other ErbBs, which would require the stripping of Hsp90 [92]. Binding of Hsp90 to ErbB2 is mediated through the loop that follows the αC-helix, and specifically within residues Gly776−Ser783 [92]. I noted that this region is located close to the asymmetric active dimer's interface, which is contributed by the N-lobe of the activated monomer (purple kinase in form 'e' in Fig. 1). Accordingly, I suggest that binding of Hsp90

prevents the formation of active dimers in which ErbB2 is the active monomer. This would lead to the preferential formation of EGFR−ErbB2 heterodimers in which EGFR is activated by ErbB2 and phosphorylates it, as reflected in the empirical results described above. It is interesting to note that alteration in the Hsp90-binding loop, as in a certain oncogenic mutant discussed in section 5.1.4, changes the normal orientation within EGFR−ErbB2 dimers. Binding of Hsp90 to the active dimerization interface also provides an explanation at the molecular level of the reluctance of ErbB2 to form homodimers.

Overall, the above findings elegantly manifest the asymmetry in the ErbBs' function within heterodimers and its importance for normal signaling. This functional asymmetry can now be clarified by the structural asymmetry of the active state observed by Kuriyan and co-workers [37].

5. Role of the ErbB Family in Pathologies

Aberrant activation of the ErbBs has been described mostly in relation to cancer [31, 93]. However, activation of the EGFR was found to be related also to kidney lesions, and ErbB4 and its ligand neuregulin-1 are involved in the pathogenesis of schizophrenia. Moreover, both EGFR and ErbB2 bind to and are activated by viruses and bacteria, which is a requirement for the pathogens' biological activity [33]. All of the above ErbB-associated pathological processes require activation of the receptors. Deficiency of ErbBs would be extremely damaging and even lethal to the developing embryo, which is presumably why it is not observed clinically.

5.1. ErbBs and Cancer

EGFR is highly expressed in many cases of non-small cell lung cancers (NSCLC) (88–99%) [94], head and neck cancers (80%) [95], and gliomas (40%), as well as in some pancreatic and breast tumors [32, 96]. Amplification and over-expression of ErbB2 have been demonstrated in 20–25% of breast cancers and also occur, albeit at lower frequencies, in lung, pancreatic, colon, endometrial and ovarian cancers [32, 97]. Accordingly, the ErbB family has become an attractive target for anticancer therapy [98].

Regardless of its frequent involvement in cancers, over-expressed EGFR is still dependent on its ligand, EGF, for activation [57, 99]; over-expression in itself is not prognostic of survival in NSCLC [94]. Correspondingly, regardless of the high percentage of EGFR involvement in NSCLC, only a small fraction of patients benefit from treatment with specific EGFR kinase inhibitors. Mutations were recently detected in the kinase domain of the EGFR

in these patients [100, 101], and proved to be oncogenic, i.e., to increase the receptor's basal kinase activity [57, 94, 99, 102, 103]. I suggest that these findings are related to the fact that random contact formation, induced by over-expression, is not by itself sufficient to provoke activation, although, statistically speaking, it probably increases the amount of active receptors within the cell. More prominent effects on activation are probably displayed by mutations that have the potential to alter the inherent activity of the receptor or to shift the dynamic equilibrium in Fig. 1 toward the active state.

Numerous mutations in ErbB family members have been detected in tumor tissues. For example, mutations in EGFR were found to occur in approximately 20% of lung cancers [104]. Mutations in ErbB2 have been reported in lung cancers, although at lower rates than EGFR mutations (1−4%) [97, 105-111], as well as in gastric, colorectal and breast cancers [109]. The mutations found in the EGFR and ErbB2 appear to be mutually exclusive, as they have never been found together in individual tumors or cell lines [104, 105, 112]. Interestingly, most of the alterations found in ErbB2 overlap with the analogous structural regions of those found in the EGFR [97], suggesting that they are functional [105]. Correspondingly, some of the prevalent mutations in EGFR and ErbB2 have been experimentally examined and found to be oncogenic (these mutations are listed in Table 1). However, the functional effect of rarer alterations (listed in Table 2) is less clear [105]. Recently, mutations in ErbB4 were also detected in NSCLC, as well as in gastric, colorectal and breast carcinomas, although at lower frequencies (1−3%) than mutations in the EGFR [111].

In the following sections I use the regulation model shown in Fig. 1A to suggest molecular interpretations, based on structural and evolutionary considerations, of the effect of known cancer-causing mutations. I use the same framework to predict the possible effects of other ErbB mutations found in cancer cells that have not been characterized experimentally (listed in Table 2).

5.1.1. Missense Mutations in the Regulatory Activation Loop

In the inactive state of the EGFR, the activation loop (residues 831−852) is packed against the αC-helix (residues 729−745) and its preceding loop (residues 723−728) such that the elements needed for catalysis are misplaced (Fig. 2). Destabilization of the packing will cause a shift toward the activated state of the kinase domain. Such an effect is indeed predicted for the oncogenic mutations L834R, which represents 41−43% of the mutations in lung cancers, and for L837Q (L858R and L861Q in pre-mature EGFR numbering, respectively) [104, 112].

L834R and L837Q are located on the activation loop and participate in hydrophobic interactions with the αC-helix in the inactive conformation (Fig. 4B), while in the active conformation they are quite exposed to the solvent (Fig. 4A). Correspondingly, mutations in these residues lead to a ligand-independent activation [57, 94, 99, 102, 103], probably via destabilization of the inactive monomeric state. In accordance with clinical findings, both EGFR mutants are more sensitive than the wt to inhibition by specific kinase inhibitors, which preferentially bind to the active conformation of the kinase domain [94, 99-103]. This further suggests that the mutants preferentially assume the active conformation. Furthermore, neither L834R nor L837Q appears to impair the maximal response to EGF [57], implying that the conformation of the mutants' kinase domain could be fully competent for catalysis.

The ligand-free state of the EGFR is characterized by interconversion between several conformations, each displaying a different binding affinity for the growth factors (Fig. 1A). Because the lower-affinity conformations are probably more energetically stable than the higher-affinity conformations, macroscopic measurements of binding affinities point to large and small populations of low- and high-affinity receptors, respectively, as discussed in section 2.5. I previously noted that alterations in EGFR can potentially affect the overall activity and ligand-binding affinity of the receptor by interference with the dynamic equilibrium depicted in Fig. 1. Mutations such as L834R and L837Q, which destabilize the inactive monomeric conformation of the kinase domain, are expected to induce monomers having ligand-independent, catalytically competent kinase domains that are sufficient to transduce signals (form a* in Fig. 4D). In addition, the L834R mutation displays approximately twofold higher affinity for EGF in what was macroscopically measured as the low-affinity binding mode [57]. Microscopically, this mutation probably shifts the dynamic equilibrium in the ligand-free state toward higher-affinity forms. Nevertheless, the EGFR mutants could not entirely mimic the ligand-bound active state because binding of the ligand further increased auto-phosphorylation of the receptor [57]. This latter event is probably related to further stabilization of active dimers by the binding of the ligand; these dimers could phosphorylate in *trans* (forms 'e−f' in Figs. 1A and 4D).

Interestingly, similarly to the EGFR, other kinases also displayed substitutions in the positions corresponding to Leu834 and Leu837 (Table 1). Overall, the high frequency of mutations in these two positions points to their crucial importance for regulation. A reasonable assumption is that EGFR alterations that destabilize the inherent inactive state of the kinase domain are likely to be oncogenic. Correspondingly, it is interesting to note that mutations in additional residues that contribute to packing of the activation loop in the

inactive state, for example I735A and M742A located on the αC-helix (Fig. 4B), show robust elevated basal activity *in vitro*, although they were not detectable in tumor cells [57].

The importance of the mutation-bearing residues is reflected also in the pattern of their substitution during evolution of the ErbB family. Leu837 is conserved as a hydrophobic residue, mostly leucine. However, similarly to the L837Q mutant found in human tumors, two oncogenic viral variants of the EGFR also display replacement by glutamine in this position, further pointing to a correlation between mutations in this residue and malignancy. Its neighbor, Leu834, is totally conserved in all catalytically active ErbBs, indicative of its important role. However, this position is substituted to valine or isoleucine in the kinase-dead ErbB3 (Val855 in human ErbB3; Fig. 4C). Similarly, the positions corresponding to Ile735 and Met742, which display other activating mutations mentioned above, are unique to ErbB3. Specifically, Met742 is conserved as methionine in all ErbBs, except for ErbB3 in some species that contain isoleucine or valine (Ile763 in human ErbB3; Fig. 4C). In addition, whereas the catalytically active ErbBs display large hydrophobic residues (such as phenylalanine, methionine, leucine, and isoleucine) in the position corresponding to Ile735, this position in the kinase-dead ErbB3 is occupied by valine, a smaller residue (Val756 in human ErbB3; Fig 4C). It is interesting to note that this position is also occupied by valine in two oncogenic variants of EGFR found in spiketail (*Xiphophorus xiphidium*} and southern platyfish (*Xiphophorus maculates),* relating this substitution to the unstable inactive state. Overall, ErbB3 displays fewer hydrophobic contacts between the αC-helix and the activation loop than the other ErbBs. For example, the contacts between Ile735 and Leu837 as well as between Met742 and Leu834 within the EGFR (Fig. 4B) are lost in ErbB3 (Fig. 4C). Together, these findings lead me to suggest that the inactive conformation of the kinase domain of ErbB3 is less stable than in the other ErbBs isoforms, and that it might even assume a constantly active conformation.

5.1.2. Mutations in the Phosphate-Binding (P)-Loop

G695S (G719S in the pre-mature EGFR numbering) is another mutation that displays heightened basal activity and is found in cancer patients [57, 94, 99, 103]. Gly695 is the first glycine in the G-X-G-X-X-G motif in the P-loop of the kinase domain, which determines the position of ATP during catalysis (Fig. 5), and accordingly is conserved in the ErbBs as well as in other tyrosine kinases. Nevertheless, the G695S mutation does not impair the maximal response to EGF, and thus allows the proper positioning of the catalytic elements to be preserved [57]. The P-loop does not display noticeable conformational changes between the

active and inactive conformations of the kinase domain. Thus, in contrast to residues in the activation loop and the αC-helix, Gly695 does not appear to participate in stabilizing the inactive conformation. The mutation might exert an effect by directly influencing the phosphate transfer reaction via lowering of the dissociation rate of ATP (as well as that of ATP-analog inhibitors). Accordingly, G695S is more sensitive than the wt EGFR to ATP-analog EGFR kinase inhibitors [99, 103]. However, because the inhibitors display higher affinity for the active conformation, G695S is probably less sensitive to the inhibitors than the L834R and L837Q mutations that induce a ligand-independent active form. Overall, I expect the G695S mutation to increase the catalytic efficiency, but not to interfere with the dynamic equilibrium in the model of Fig. 1. Experimental results indeed show that G695S does not alter the binding affinity for the ligand [57].

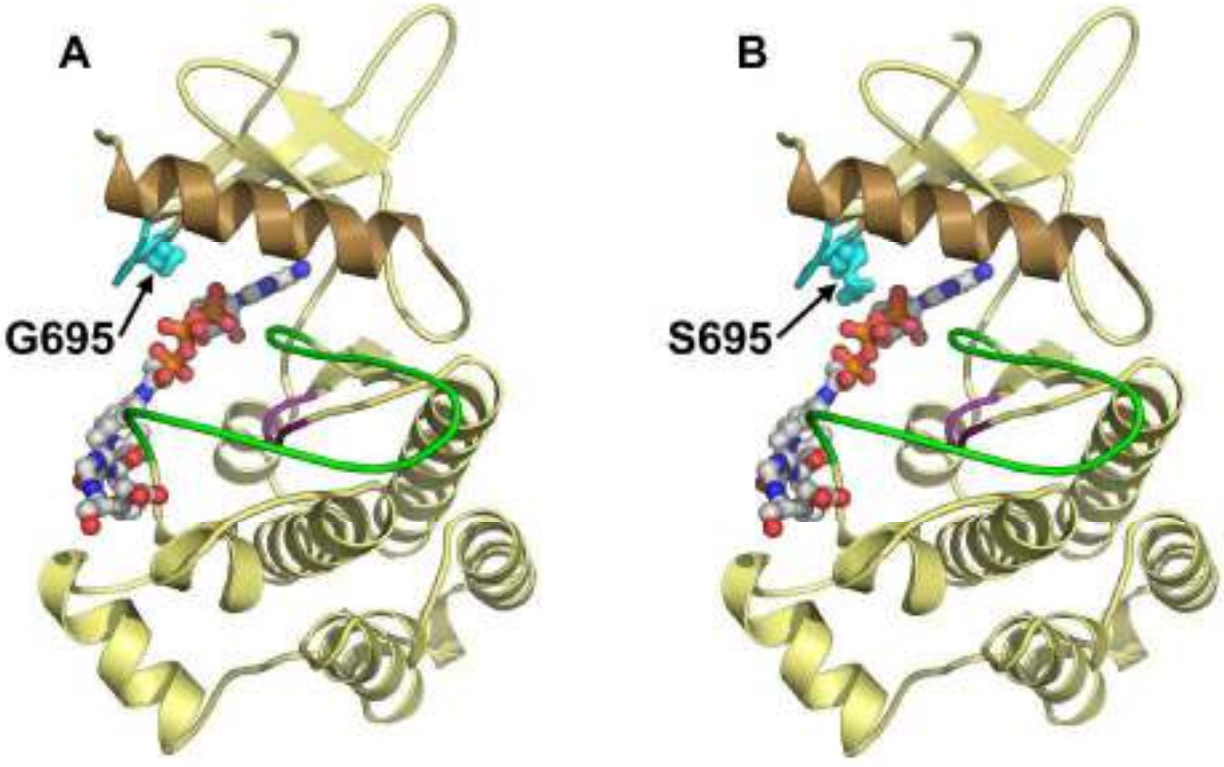


Fig. 5: Mutations in the Phosphate-Binding Loop

The kinase domain of the EGFR (residues 685−957) is depicted in a yellow ribbon representation. Regulatory elements are depicted by different colors: the activation loop (residues 831−852) in green, the αC-helix (residues 729−745) in brown, the phosphate-binding loop (residues 695−700) in cyan, and the catalytic loop (residues 812−818) in purple. Position 695 is displayed in a balls-and-sticks representation. **(A)** The wt EGFR (PDB entry – 2gs6 ; [37]). **(B)** Model of the G695S mutant found in cancer cells.

5.1.3. Mutations at the Interface of the Active Asymmetric Dimer

Substitutions of alanine or glycine for Glu685 (709 in the pre-mature EGFR numbering) have been reported in lung cancers [113, 114]. Like the other missense mutations, E685G also shows increased basal auto-phosphorylation and a heightened sensitivity to EGFR kinase inhibitors such as gefitinib [94]. Interestingly, Glu685 is located on the N-lobe of the EGFR as part of the interface on the activated monomer within the asymmetric dimer (purple kinase in Fig. 6). This interface is important for activation of the catalytically active ErbBs, but is irrelevant for the kinase-dead ErbB3 [37], as discussed in section 4.3.3. Accordingly, Glu685 is conserved within the catalytically active ErbBs, but not in some species of ErbB3. I suggest that the E685G mutation might further stabilize the formation of the active dimer by shifting the dynamic equilibrium model of EGFR towards the active forms (e.g. form 'c' in Fig. 1A), leading to an increase in basal activity. Another possibility is that within heterodimers comprising EGFR, mutations at the interface might change the native orientation of the monomers by altering the roles of the activating and the activated monomers (see section 4). The resulting modification in signal transduction could also lead to oncogenic transformation (for an example see section 5.1.4).

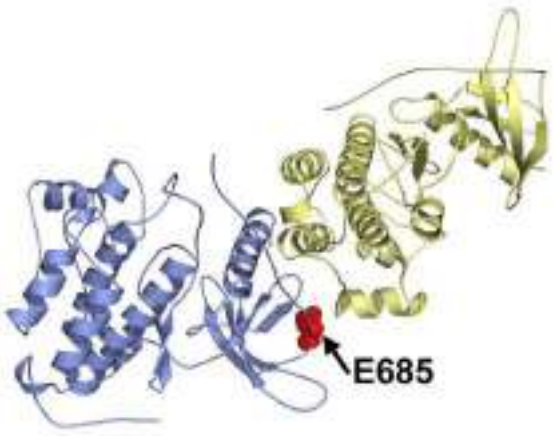


Fig. 6: Mutation at the Interface of the Intracellular Active Dimer

The asymmetric active dimer of the kinase domain (colored yellow and purple) of the EGFR (residues 685−967; PDB entry – 1m17; [38]) is depicted in a ribbon representation. Glu685, located on the activated monomer (in light purple), is depicted as a red sphere. This residue is substituted in some cancer cells.

5.1.4. Mutation in the αC-helix and its Surrounding Regions

EGFR Exon19 *deletions in the N-terminus of the αC-helix and its preceding loop*

In addition to the missense mutations discussed above, other oncogenic variations in the EGFR correspond to different deletions within residues 722−735 in *exon19*. As with the former mutations, when tested *in vitro* such deletions show increased basal activity compared to wt EGFR [57, 94, 99, 102], as well as higher sensitivity to EGFR kinase inhibitors [94, 99,

100, 102]. The region displaying the deletions constitutes the loop preceding the αC-helix and its N-terminus, which participate in interactions with the activation loop in the inactive conformation (Fig. 7A). The crystal structure of the inactive conformation of the kinase domain [37] indeed displays pronounced rigidity in this region, in contrast to the marked flexibility observed in the active conformation [37, 38]. Thus, such deletions are predicted to lead to destabilization of the inactive conformation of the kinase domain and to an increase in basal activity, similarly to the L834R and L837Q mutations discussed in section 5.1.1.

I note that although the deletions destabilize inhibitory interactions, the maximal catalytic activity was diminished [57], probably because the αC-helix is involved in mediating the active intracellular dimer (Fig. 7C). Accordingly, the ΔL723-P729insS mutant (a 723−729 deletion together with insertion of a serine residue) was shown to abolish the high-affinity binding mode and reduce the low-affinity binding mode by twofold [57] (see section 2.5 for discussion of ligand-binding affinity). This corresponds to destabilization of the active intracellular dimer constituting the high-affinity forms (forms 'c−d' in Fig. 1A), and results in a general shift toward lower-affinity states.

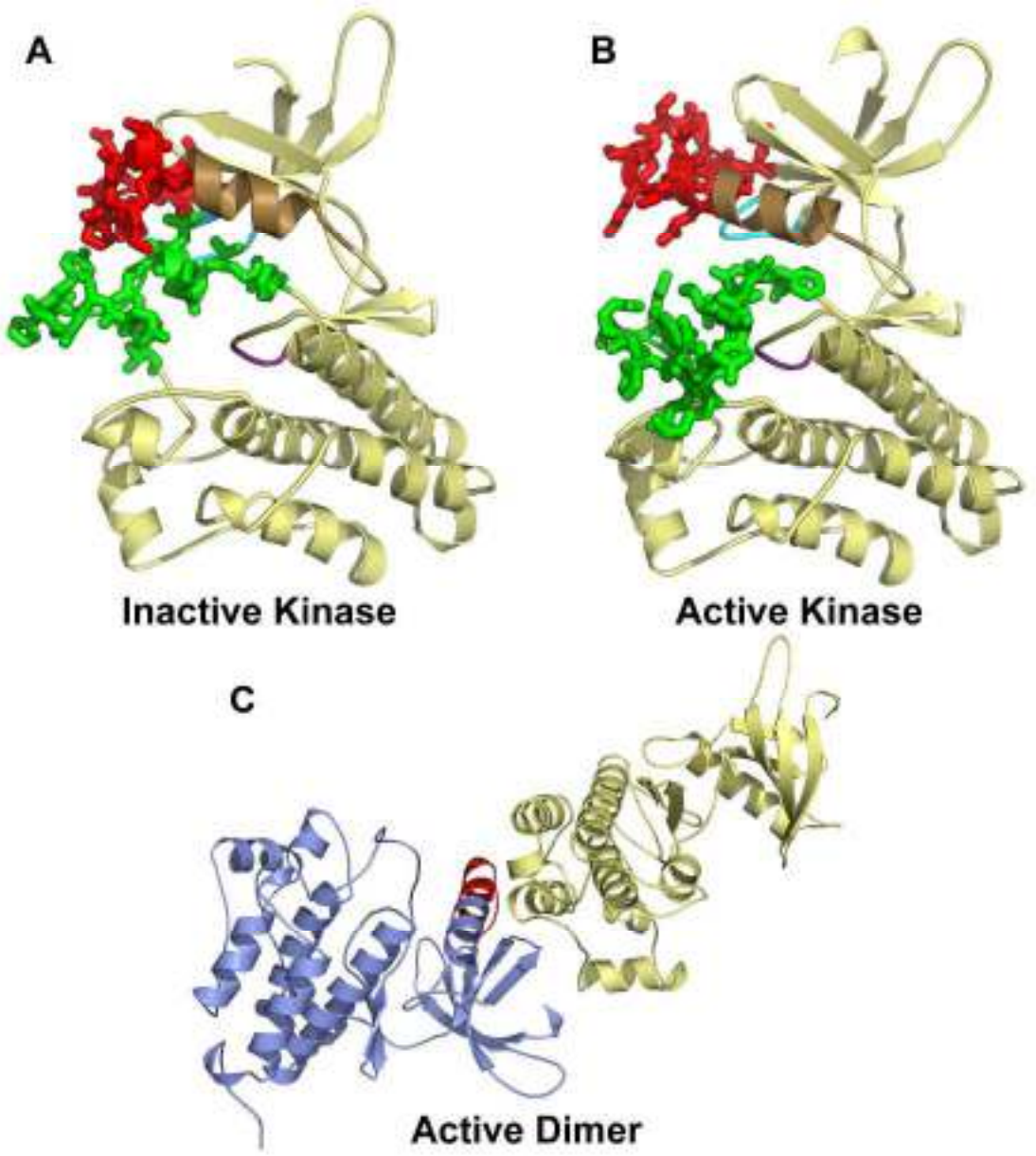


Fig. 7: Deletions in *Exon19* in the EGFR

(A−B) The kinase domain of the EGFR (residues 685−957), in the inactive **(A)** (PDB entry – 1gs7; [37]) and active **(B)** (PDB entry – 2gs6 ; [37]) conformations, is depicted in a yellow ribbon representation. Regulatory elements are depicted by different colors: the activation loop (residues 831−852) in green, the C-terminus of the αC-helix (residues 736−745) in brown, the phosphate-binding loop (residues 695−700) in cyan, and the catalytic loop (residues 812−818) in purple. The region showing deletions in cancer cells (residues 722−735; displayed in a balls-and-sticks representation) is colored red. Residues comprising the activation loop are also displayed in a balls-and-sticks representation. Interactions between the region showing deletions in cancer cells (red) and the activation loop (green) in the inactive state (A) are noticeable. **(C)** The asymmetric dimer of the kinase domains (the monomers are colored yellow and light purple) of the EGFR (residues 685−957; PDB entry- 1m17; [38]) is depicted in a ribbon representation. The region showing deletion in cancer cells (residues 722−735) is colored red in the activated (purple) monomer.

Insertions of exon 20 *in the loop following the αC-helix*

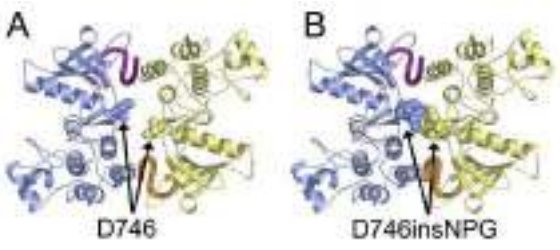
Exon 20 insertions have been detected both in the EGFR and in ErbB2, and were shown to be activating alterations. They are discussed in the following sections.

Exon20 insertions within the EGFR destabilize the inactive dimer

In contrast to the EGFR alterations discussed above, the *exon20* insertion mutations have not been reported in the group of clinical responders to EGFR kinase inhibitors. Nevertheless, they were detected in few recent large-scale studies [94, 95, 113, 114], and reportedly account for 3.7−5% of the mutations in lung cancers [104, 112]. The insertions were detected in the C-terminus of the αC-helix and the region following it, within residues 744−749. One such insertion, D746insNPG, i.e., insertion of residues Asn-Pro-Gly following Asp746 (Asp770 in pre-mature EGFR numbering), was tested *in vitro* and showed heightened basal activity [99]. In contrast to the previously discussed mutations and in accordance with clinical findings, this insertion mutant was remarkably insensitive to the EGFR kinase inhibitors gefitinib and erlotinib; it was more sensitive, however, to treatment with the irreversible inhibitor CL-387,785 [99], which is covalently bound to EGFR in residue C773 [115].

The region following the αC-helix forms an exposed loop on the kinase domain. This loop displays similar orientations in the active and inactive states, and is not predicted to play a role in stabilizing the monomeric inactive state. Nevertheless, it is located close to the interface of the symmetric, putatively inactive, crystallographic dimer of the kinase domain (see section 2.2.1), facing the equivalent loop from the second monomer (Fig. 8). My prediction of the structure of the D746insNPG alteration showed clashes between insertions within the symmetric dimer (Fig. 8), possibly leading to destabilization of the inactive form (form 'b' in Fig. 1), and prompting activation by increasing the fraction of receptors in the active form. The mechanism leading to insensitivity to the inhibitors is not yet clear.

Fig. 8: Insertions in *Exon20* in the EGFR



The intracellular symmetric inactive dimer (PDB entry – 1m17; [38]) is depicted in a ribbon representation. The two kinase domains are colored yellow and light purple and their following C-terminal fragments are depicted as orange and purple tubes, respectively. **(A)** wt EGFR; Asp746 is depicted in space-filled atoms. **(B)** Model of the EGFR D746insNPG mutant. Residues comprising this insertion are depicted as space-filled atoms. Clashes within the inactive dimer, resulting from the insertion, are noticeable.

Exon20 insertions in ErbB2 alter orientation within EGFR−ErbB2 heterodimers

The most prevalent ErbB2 alterations found to date in cancer cells are insertions in *exon20*. These insertions were found within residues 774−783 of ErbB2, corresponding to residues 742−751 of the EGFR [95, 97, 105-107, 110, 116]. The most common of these alterations, G776insYVMA, was recently shown to exhibit a more potent auto-catalytic activity than wt ErbB2 [91]. Interestingly, this insertion is located within the binding site of Hsp90 [92].

In section 4.3 I discussed functional asymmetry within EGFR−ErbB2 heterodimers, wherein ErbB2 is preferentially the activating monomer and makes contact with the EGFR via its C-lobe (the yellow kinase in form 'e' in Fig. 1). The EGFR in turn takes on the role of activated monomer and phosphorylates ErbB2, leading to specific signal transduction. This orientation is probably imposed by binding of Hsp90 to the dimerization interface on the N-lobe of ErbB2 [92] (see section 4.3.2). Alterations in the Hsp90 binding loop, as in the G776insYVMA mutant, probably prevent the binding of Hsp90 and enable ErbB2 to interact with the EGFR via its N-lobe. As a result, ErbB2 preferentially assumes the role of the activated monomer (the purple kinase in form 'e' in Fig. 1). Correspondingly, experiments have shown that in contrast to wt ErbB2, the G776insYVMA mutant phosphorylates its heterodimeric partner EGFR in a manner that does not require a catalytically active EGFR. Accordingly, in contrast to the situation in wt dimers, this activation was blocked by direct ErbB2 inhibitors such as trastuzumab, lapatinib, and CI-1033, but not by the EGFR-specific kinase inhibitors erlotinib or gefitinib [91]. In addition, binding of Hsp90 restrains cellular signaling by ErbB2 [92], indicating that the kinase activity of the insertion mutant is more potent than that of the wt [91].

I note that an additional contribution prompting formation of the active dimer by the mutation might be related to destabilization of the inactive dimer, as in the case of *exon20* insertions within the EGFR. Interestingly, heterodimers constituting the EGFR and the G776insYVMA ErbB2 mutant do not respond to ligand binding, and do not require contact formation between their extracellular domains [91]. It seems likely that the association of ErbB2 via its N-lobe to the EGFR's C-lobe is energetically stable and that, together with a probable loss of control by the inactive dimer, it could preferentially occur via random contacts. Under normal conditions, this interaction is prevented mainly by binding of Hsp90 to the dimerization interface and the formation of an inactive dimer.

The S744I mutation at the C-terminus of the αC-helix

Another residue that undergoes substitution in tumor cells is Ser744, located at the C-terminal end of the αC-helix. Compared to the wt, the S744I mutant displays heightened basal kinase activity and greater sensitivity to EGFR kinase inhibitors such as gefitinib [94]. In contrast to mutations in the activation loop and the αC-helix, which are buried inside the hydrophobic core and thereby stabilize the inactive conformation, Ser744 is relatively exposed to the solvent. The effect, therefore, is probably not related to destabilization of the inherent inactive conformation.

Ser744 is conserved in ErbBs except for ErbB2, in which a glycine residue occupies this position (Gly776 in human ErbB2). This region in ErbB2 indeed displays the unique ability, found in the ErbBs, to bind Hsp90 [92]. Interestingly, an equivalent G776S mutation in ErbB2 was found in gastric tumors [97]. The equivalent locations of the EGFR-S774G and ErbB2-G776S mutations might point to a common molecular effect. Because this position is close to the interfaces of the active dimer, the mutation might contribute to alteration of the native orientation within active asymmetric heterodimers, as in the ErbB2 *exon20* insertions discussed above.

5.1.5. Mutations Found in Tumor Cells that have not been yet analyzed *In Vitro*

A large set of mutations has been found to exist in tumor cells. Those occurring more frequently have been analyzed *in vitro* for their effects on kinase activity, as discussed in sections 5.1.1−5.1.4. The oncogenic nature of other mutations of the EGFR, ErbB2, and ErbB4 has not yet been established. Based on the structural locations of these residues and their evolutionary conservation patterns within the ErbB family, I offer (Table 2) a prediction as to the nature of their substitutions, i.e., whether it is likely to be damaging or neutral. Some examples are provided below.

The R752C mutation in the EGFR is predicted to be damaging

R752C (residue 776 in the pre-mature EGFR numbering) was found as a second mutation to L834R in patients with lung tumors sensitive to the EGFR kinase inhibitor erlotinib [101]. Arg752 is conserved in ErbBs and is located at the interface of the symmetric, putatively inactive crystallographic dimer (Fig. 3). This residue, which is connected by a salt bridge to Asp990 located on the C-terminal domain, plays a role in the polar network stabilizing the dimer (see section 2.2.1). Thus, this mutation is predicted to lead to destabilization of the inactive dimeric state. Interestingly, a mutation in the corresponding residue of ErbB4

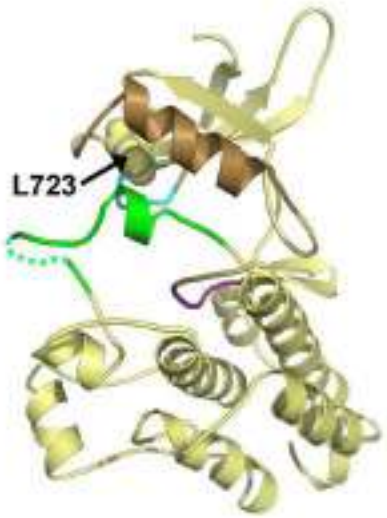
(R782Q) [111] was also reported in cancer cells, supporting my assumption that these mutations are indeed oncogenic.

The L723F mutation in the EGFR is predicted to be damaging

The L723F mutation *(*residue 747 in the pre-mature EGFR numbering) is observed in lung cancers [117]. Residue Leu723 is located in the loop preceding the αC-helix, at the beginning of the *exon19* deletions found in cancer patients, and is in contact with the activation loop in the inactive state, i.e., with residues Leu834, Leu837, and Leu838 (Fig. 9). Leu723 is conserved in all the ErbBs except the kinase-dead ErbB3, in which this position is occupied by isoleucine, and which probably exhibits a constant active conformation of the kinase domain (see section 5.1.1). In view of its conservation pattern and structural location, I predict that the Leu723 mutation might participate in destabilizing the inactive state of the kinase domain, similarly to the *exon19* deletions and other missense mutations in the αC-helix and activation loop. Interestingly, the corresponding residue in ErbB2 (Leu755) also displays substitutions in cancer cells [97, 107, 110], suggesting that these mutations are indeed oncogenic.

Fig. 9: A Predicted Oncogenic Mutation

The kinase domain of the EGFR (residues 685−957) in the inactive conformation (PDB entry – 1gs7; [37]) is depicted in a yellow ribbon representation. Regulatory elements are depicted by different colors: the activation loop (residues 831−852) in green, the αC-helix (residues 729−745) in brown, the phosphate-binding loop (residues 695−700) in cyan, and the catalytic loop (residues 812−818) in purple. The disordered region in the crystal structures is depicted by a dashed line. Leu723, which was found to be mutated in cancer cells, is depicted in space-filled atoms. In this inactive conformation Leu723 interacts with Leu834, Leu837, and Leu838 on the activation loop (depicted in green space-filled atoms). I predict that the L723F substitution leads to destabilization of the inactive conformation.



The N857S mutation in ErbB2 is predicted to be neutral

The ErbB2 N857S mutant (corresponding to Gln825 in EGFR) was reported in an ovarian tumor [97]. A model of ErbB2 predicted that Asn857 is located in a loop on the back side of the catalytic site and is exposed to the solvent. This position shows evolutionary variation in

ErbBs. Based on analyses of evolutionary conservation and structural location, I predict that substitution of the mutation N857S is neutral in its effect.

6. Methods

6.1. Evolutionary Conservation Analysis of the ErbB Family

Sequences of the kinase domain of ErbB isoforms from various species were collected from the UNIPROT database [118] and the NR database from NCBI, using PSI-BLAST [119]. The resulting 306 sequences were aligned using MUSCLE [120]. Fragmented or redundant sequences were removed from the alignment. In addition, sequence variants and mutants were discarded, along with proteins sequenced by a whole-genome shotgun project, because they are viewed as preliminary data. The resulting multiple sequence alignment (MSA) contained 48 sequences of ErbBs from vertebrates and invertebrates. The invertebrate sequences were removed, and an MSA of 36 ErbBs from vertebrates (and viruses) was used to calculated evolutionary conservation scores using a Bayesian method [121] as implemented in the ConSurf web-server (http://consurf.tau.ac.il/) [5].

6.2. Structure Prediction and Analyses

The structures of the kinase domains of ErbB2, ErbB3, and mutants found in cancer cells were modeled on the basis of the structure of the kinase domain of the EGFR (PDB entry – 1m17; [38]) using the NEST program [22] with default parameters. The solvent-accessible area was calculated using the SURFV program with a probe sphere of radius 1.4 Å and default parameters [9].

7. Tables

7.1. Table 1: Oncogenic mutations in EGFR and ErbB2, found in cancer cells and characterized in vitro

Mutation	Cited Report	Experimental Results	Functional and Structural Characterization	Notes
EGFR Alterations				
Missense Mutations in the EGFR				
E685X (E709X) X = [A,G,K,H]	Mutations in Glu685 were reported in lung cancer [113, 117]. E685H was found as a second mutation to Leu834 or to Gly695 [114].	Mutations in Glu685 increased the basal activity of the EGFR and its sensitivity to EGFR kinase inhibitors [94].	Glu685 is located on the interface of the active asymmetric dimer, and is predicted to stabilize a specific active dimeric form or change the native interactions between the activated and activating monomers within the dimer.	
G695X (G719X) X = (A, C, S, D)	Mutations in Gly695 were reported in lung cancer [95, 100, 117, 122].	Mutations in Gly695 increased the basal and EGF-induced activity of the EGFR, its oncogenic potential at the cellular level, and its sensitivity to EGFR kinase inhibitors [57, 94, 99, 102, 103]. Mutations showed no evidence of impairing the maximal EGF response or altering the binding affinity for EGF [57]. We note that the effects were seen only with stable	Gly695 is the first glycine within the GXGXXG motif in the phosphate-binding loop (P-loop). The mutation might exert its effect by directly influencing the phosphate transfer reaction via lowering the dissociation rates of ATP (and of ATP-analog inhibitors). Accordingly, G695S is more sensitive than the wt EGFR to ATP-analog EGFR kinase inhibitors [99, 103]. However, probably because the inhibitors display higher affinity for the active	Mutations in Gly695 represent 3%−4% of mutations in lung cancers [104, 112]. Of patient who sustained this mutation, 56% responded to treatment with EGFR kinase inhibitor [104]. A mutation in the corresponding residue in the

		expressions of the mutant.	conformation, G695S is less sensitive to the inhibitors than the L834R and L837Q mutations that induce a ligand-independent active form.	B-raf Ser/Thr kinase (Gly463) was detected in cancer cells and was oncogenic in vitro [123].
S744I (S768I)	S744I was reported in lung cancer [95, 101]. S744I was also reported as a second mutation to G719C/S [113].	Compared to the wt, S744I displayed increased basal kinase activity and greater sensitivity to EGFR kinase inhibitors such as gefitinib [94].	Ser744 is located at the C-terminus of the αC-helix. Because this position is close to the interface of the active dimer, the effect of the mutation may be related to alteration in the native orientation within active asymmetric hetero-dimers.	The corresponding residue in ErbB2 (Gly766) also undergoes substitutions in cancer cells [97], suggesting that these mutations are indeed oncogenic.
L834X (L858X) X = [R,M]	L834R was reported in lung cancer [95, 100, 101, 106, 113, 117, 122]. L834M was reported in lung cancer [117].	L834R increased basal and EGF-induced activity of the EGFR, its oncogenic potential at the cellular level, and its sensitivity to EGFR kinase inhibitors [57, 94, 99, 102, 103]. L834R did not appear to impair the maximal EGF response. However, its affinity for EGF in the low-affinity binding mode was twofold greater than that of the wt [57]. We note that the effects were seen only with stable expressions of the mutants.	L834R is located on the activation loop. It is predicted to cause destabilization of the inactive conformation of the kinase domain.	L834R accounts for 41%−43% of the mutations in lung cancers [104, 112]. Of the patient who sustained this mutation, 71% responded to treatment with EGFR kinase inhibitor [104]. A mutation in the corresponding residue in the B-raf Ser/Thr kinase (Leu596) was detected in cancer cells and was oncogenic in vitro [123].

L837Q (L861Q)	L837Q was reported in lung cancer [100, 106, 113, 117].	L837Q increased the basal and EGF-induced activity of the EGFR, its oncogenic potential at the cellular level, and its sensitivity to EGFR kinase inhibitors [57, 94]. L837Q did not appear to impair the maximal EGF response [57].	L837Q is located on the activation loop. It is predicted to cause destabilization of the inactive conformation of the kinase domain.	The corresponding residue in ErbB2 (Leu869) also undergoes substitutions in cancer cells [109], suggesting that this mutations are indeed oncogenic. Also, mutations in the corresponding positions in the B-raf Ser/Thr kinase (Val599), the murine (D814V) [124] and human (D816V/H) [125] c-Kit tyrosine kinase, and the C-fms tyrosine kinase (D802V) [126] were detected in cancer cells and were oncogenic in-vitro [123, 124, 126].

In-Frame Deletions in *Exon19* in the EGFR				
ΔE722−A726 (ΔE746−A750)	Reported in lung cancer [95, 100, 101, 113, 114, 122].	The deletions increased the basal and EGF-induced activity of the EGFR, its oncogenic potential at the cellular level, and its sensitivity to EGFR kinase inhibitors [57, 94, 99, 102]. ΔL723-P729insS and ΔS728-I735 showed substantially reduced maximal levels of EGF-induced auto-phosphorylation [57]. We note that the effects were seen only with stable expressions of the mutants.	The deletions in *exon19* are located in the loop preceding the αC-helix and its N-terminus. They are predicted to cause destabilization of the inactive conformation of the kinase domain. However, the deletions would also compromise both the catalytically competent conformation of the kinase domain and the interactions within the active intracellular dimer, leading to reduced ligand-induced activity.	*Exon19* deletion accounts for 44%−48% of mutations in lung cancers [104, 112]. Of the patient sustaining these deletions, 84% responded to treatment with EGFR kinase inhibitor [104].
ΔL723−S728 (ΔL747−S752)	Reported in lung cancer [101].			
ΔL723−P729insS (ΔL747−P753insS)	Reported in lung cancer [100].			
ΔL723−E725_A726P (ΔL747−E749_A750P)	Reported in lung cancer [99].			
ΔS728−I735 (ΔS752−I759)	Reported in lung cancer [114, 122].			
In-Frame Insertions/Duplications in *Exon20* in the EGFR				
D746insNPG (D770insNPG)	Reported in lung cancer [94].	The insertion increased the basal activity of the EGFR. This mutant was remarkably insensitive to gefitinib and erlotinib. Consistent with this result, all three lung	The insertions in *exon20* are located in the loop following the αC-helix. This loop is located close to the interface of the symmetric, putatively inactive crystallographic dimer of the kinase	Overall *exon20* insertions account for 3.7%−5% of the alterations in lung cancers [104, 112].

		adenocarcinoma patients with known *exon20* insertion mutants of EGFR failed to show a clinical response to treatment. However, the mutant was sensitive to treatment with an irreversible inhibitor, CL-387,785[99].	domain, facing the equivalent loop from the second monomer. The insertions are predicted to cause destabilization of the inactive form. The mechanism causing insensitivity to the inhibitors is not yet clear.	
		ErbB2 Alterations		
		In-frame Insertions/Duplications in *Exon20* in ErbB2		
G776insYVMA (Ser744 in the EGFR).	Insertions following residue Gly776 were reported in cancer cells [91, 105, 107, 110, 116].	G776insYVMA showed more potent autocatalytic activity than wt ErbB2 [91]. Heterodimers constituting the EGFR and the G776insYVMA ErbB2 mutant did not respond to ligand binding; moreover, they did not require contact formation between their extracellular domains [91].	These insertions are located within the binding site of Hsp90 [92], and close to the interface of the asymmetric active dimer of the kinase domain. Binding of Hsp90 restrains the activity of ErbB2, probably because of interference with formation of the asymmetric active dimer in which ErbB2 contacts the second monomer via its N-lobe. The alterations probably affect the active dimeric state. Specifically, they alter the native orientation of the monomers within the heterodimer such that, in contrast to the wt dimer, ErbB2 is the	The most prevalent ErbB2 alterations found to date in cancer cells are insertions in *exon20*. These insertions were detected within residues 774−783 of ErbB2, which correspond to residues 742−751 of the EGFR [95, 97, 105-107, 110, 116].

			activated monomer that phosphorylates the EGFR.	

Table 1: Alterations in EGFR are classified into three groups: missense mutations, *exon19* deletions, and *exon20* insertions. Alterations in ErbB2 comprise only an insertion in *exon20*. Positions of the mutations in the mature EGFR are indicated (numbering in pre-mature EGFR is in parenthesis). For the mutation in ErbB2, the corresponding residue in the mature EGFR is indicated in parenthesis.

7.2. Table 2: Mutations in EGFR, ErbB2, and ErbB4, detected in cancer cells but not yet characterized in-vitro.

Mutation	Cited Report	Functional and Structural Characterization	Predicted Effects
EGFR Alterations			
Missense Mutations in the EGFR			
G700S (G724S)	Reported in lung cancer [117].	Gly700 is located on the P-loop within the glycine-rich motif. It is conserved in ErbB and tyrosine kinases in general.	G700S is predicted to display a similar effect to that of the G695X mutation, which increases catalytic activity. The mutation might enhance catalysis. We note that the corresponding position in the B-Raf Ser/Thr kinase (G468A) was also found to be mutated in cancer cells [123].
E710K (E734K)	Reported in lung cancer [117].	Glu710 is located close to the interface within the active asymmetric dimer contributed by the N-lobe. This residue could be involved in a salt-bridge with the second monomer. This position is conserved as aspartate or glutamate residues in ErbBs from vertebrates.	E710K could have an effect on the active dimeric state.
L723F (L747F)	Reported in lung cancer [117].	Leu723 is located at the loop preceding the αC-helix, at the beginning of many *exon19* deletions found in cancer patients. It forms contacts with the activation loop in the inactive state and with the αC-helix in the active	The mutation might be involved in destabilization of the inactive state of the kinase domain, like the *exon19* deletions and other missense mutations in the αC-helix and activation loop. We note that the

		state. Leu723 is conserved in ErbBs from vertebrates, except for ErbB3, which displays isoleucine in this position.	corresponding residue in ErbB2 (Leu755) also undergoes substitutions in cancer, suggesting that these mutations are indeed oncogenic. The exceptional residue occupying this position in ErbB3 might point to a less stable inactive state of the kinase domain in ErbB3.
R724P (R748P)	Reported in lung cancer by [117].	Arg724 is located on the loop preceding the αC-helix. However, its side chain faces the solvent and does not participate in the inactive hydrophobic packing of the kinase domain. It is not conserved in the ErbBs, not even within the same ErbBs in different species.	Predicted to be polymorphic, according to the evolutionary conservation analysis and structural location.
V745X (V769X) X = [M, L]	V745L was reported together with S744I (S768I) on the same sequence [114]. V745M was reported together with an in-frame deletion in *exon19* on the same sequence [113].	Val745 is located at the end of the αC-helix, but does not participate in the packing with the activation loop in the inactive state. Conserved as a hydrophobic residue within the ErbBs, mostly as valine, leucine, and few methionine residues.	Because this residue is located in a regulatory element the mutation might be damaging, although the mild nature of the substations weakens this prediction. Nevertheless, the fact that the corresponding residue in ErbB2 (Val777) undergoes similar substitutions in cancer cells points to the oncogenic nature of these mutations.
H749R (H773R)	H749R was reported together with W731Stop in patients who did not respond to EGFR kinase inhibitors. [113].	The position corresponding to H749R in ErbBs from vertebrates is occupied by histidine or tyrosine. It is also conserved in tyrosine kinases as histidine, tyrosine, or asparagine residues. It is substituted for	Because of its conservation pattern in tyrosine kinases and ErbBs, and its close proximity to an essential residue, this mutation is predicted to be damaging. However, the nature of the effect is not yet known.

		arginine in an oncogenic viral EGFR variant. His749 is located spatially close to Lys799, which was shown to be essential for EGFR function [37].	
R752C (R776C)	R752C was reported as a second mutation to L834 or G695X [101, 114].	Arg752 is located on the interface of the symmetric, putatively inactive dimer. It is conserved in ErbBs from vertebrates.	R752C might participate in destabilization of the inactive dimeric conformation, leading to heightened basal activity. We note that the corresponding residue in ErbB4 also undergoes substitution in cancer cells, suggesting that these mutations are indeed oncogenic.
Q763R (Q787R)	Reported in lung cancer [117].	Gln763 is conserved in ErbBs from vertebrates and is buried within the protein in both the active and the inactive conformations.	Because of its conservation pattern in ErbBs and its structural location, Q763R is predicted to be damaging. However, the nature of the effect is not yet known.
T766M (T790M)	Reported in lung cancer [117].	Thr766 is located in the ATP-binding pocket, and contacts the ATP-analog in the crystal structure [38]. It is conserved in ErbBs from vertebrates.	T766M is responsible for at least half of the acquired resistance to EGFR kinase inhibitors such as gefitinib and erlotinib. Substitution of a bulkier residue, such as methionine, for threonine is thought to sterically hinder the binding of these drugs (reviewed in [104]). A mutation in the corresponding residue in the ABL1 tyrosine kinase (T315I) is also related to an acquired resistance in ABL1, (reviewed in [104]).

L809V (L833V)	L809V was reported in lung cancer [117]. L809V was also reported as a second mutation to H811L [113].	Leu809 is located close to the activation loop and αC-helix in the inactive conformation, and contacts Phe832 from the conserved DFG motif [127]. Leu809 is conserved as a leucine residue in all ErbBs except ErbB3, which displays methionine in this position. A methionine in this position is also displayed by a viral variant of the EGFR.	L809V might participate in destabilization of the hydrophobic packing in the inactive conformation of the kinase domain. The exceptional residue occupying this position in ErbB3 might point to a less stable inactive state of the kinase domain in ErbB3.
V810L (V834L)	V810L was reported in lung cancer [117].	Val810 interacts with residues from the conserved catalytic loop (residues 811−813). In the inactive conformation it is packed directly against Arg812. Val810 is largely conserved in ErbBs.	The mutation might stabilize the active state of the catalytic loop in the kinase domain. We note that the corresponding residue in ErbB2 (Val842) also undergoes substitution in cancer cells, suggesting that these mutations are indeed oncogenic.
H811L (H835L)	H811L was reported in lung cancer [101].	His811 is a part of the conserved catalytic loop (residues 811−813; Asp813 is the proton acceptor catalytic residue). His811 is totally conserved in ErbBs and tyrosine kinases.	The evolutionary conservation and location of His811 suggest that mutation in this position would be damaging. However, the nature of the effect is not yet known.
L814V (L838V)	L814V was reported in lung cancer as a second mutation to L834R [113].	Leu814 follows the conserved catalytic loop (residues 811−813). Leu814 is conserved in ErbBs and is completely buried within the protein.	Because of the conservation pattern and structural location of this residue, the mutation is predicted to be damaging. However, the nature of the effect is not yet known.
A815T (A839T)	A815T was reported in lung cancer [113].	Ala815 follows the conserved catalytic loop (residues 811−813).	Because of the conservation pattern and structural location of this residue, the mutation is

		Ala815 is conserved in ErbBs and is completely buried within the protein.	predicted to be damaging. However, the nature of the effect is not yet known.
K822R (K846R)	K822R was reported in lung cancer [113].	Lys822 is located on the interface of the crystallographic symmetric dimer involved in the polar network within the complex. Lys822 is conserved in ErbBs from vertebrates.	K822R might lead to destabilization of the inactive dimer.
G849E (G873E)	G849E was reported in lung cancer [117].	Gly849 is located on the activation loop. Gly849 is conserved in ErbBs from vertebrates, except for ErbB3, which displays a glutamate residue in this position (as in the mutant).	G849E might lead to destabilization of the inactive state of the kinase domain. Glutamate was previously shown to mimic a phosphate in the activation loop, leading to activation.
In-Frame Deletions in *Exon19* in the EGFR			
Δ722−727 insA (Δ746−751insA)	Reported with V769M [113].	The *exon19* deletions are located in the loop preceding the αC-helix and in its N-terminus.	*Exon19* deletion accounts for 44%−48% of mutations in lung cancers [104, 112]. Of the patient sustaining these deletions, 84% respond to treatment with EGFR kinase inhibitors [104]. These deletions are predicted to have a similar effect to those tested in vitro (such as ΔL723-P729insS), which show increased basal activity and enhanced transforming potential [57], probably because of destabilization of the inactive conformation of the kinase domain.
Δ722−727 insI (Δ746−751insI)	Reported by [101].		
Δ722−728 insV (Δ746−752insV)	Reported by [95].		
Δ722−728 insD (Δ746−752insD)	Reported by [113].		
Δ723−726insP (Δ747−750insP)	Reported by [114].		
Δ723−727 (Δ747−751)	Reported by [113, 114].		
Δ723−727insS (Δ747−751insS)	Reported by [100].		
Δ723−728 (Δ747−752)	Reported by [101, 114].		

Δ723−728 insQ (Δ747−752insQ)	Reported by [101].		
Δ723−729insS (Δ747−753insS)	Reported by [100].		
720insKIPVAI (744insKIPVAI)	Reported by [114].	Located on a β-strand in the N-lobe.	This is an unexpected insertion in *exon19*. The effect of this mutation could not be predicted.
In-Frame Insertions/Duplications in *Exon20* in the EGFR			
D737insEAFQ (D761insEAFQ)	Reported by [113, 114].	The *exon20* insertions are located in the C-terminus of the αC-helix and its following loop.	*Exon20* insertions constitute 3.7%−5% of alterations in lung cancers [104, 112]. *Exon20* insertions, like the wt EGFR, are sensitive to EGFR kinase inhibitors [104]. All of these insertion are predicted to show increased basal activity and enhanced transforming potential, similar to the D746insNPG alteration tested in vitro [99]. Nevertheless, the molecular effects of these insertions might differ, depending on their exact location. For example, insertions C-terminal to residue 746 are predicted to interfere with the interface of the symmetric, putatively inactive dimer. On the other hand, insertions located at the C-terminus of the αC-helix might participate in destabilization of the inactive state of the kinase domain. Alternatively, because the αC-helix constitutes the interface of the active dimer, these insertions might have an effect dimer formation.
A743insTLA (A767insTLA)	Reported by [114].		
Dup744-746 (Dup768−770)	Reported by [95].		
V745insASV (769insASV)	Reported by [114].		
D746insNPG (D770insNPG)	Reported by [94].		
D746GinsY (D770GinsY)	Reported by [114].		
Dup747−749 (Dup771−773)	Reported by [95].		

ErbB2 Alterations			
Missence Mutations in ErbB2			
K724N (Lys692 in the EGFR)	K724N was reported in gastric cancers [109].	A model of ErbB2, based on the EGFR, predicted that Lys724 is exposed to the surface and is located close to the interface of the symmetric putatively inactive dimer. Lys724 is conserved in ErbBs from vertebrates.	Lys724 might be important in the electrostatic complementarity between the kinase and the C-terminal domains within the symmetric dimer. Therefore, the mutation might lead to destabilization of the inactive dimeric state.
T733I (Thr701 in the EGFR)	T733I was reported in gastric cancers [109].	Thr733 is located within the glycine-rich motif on the P-loop. It is conserved in ErbBs from vertebrates.	T733I might affect the phosphate transfer reaction.
L755X X= [P, S] (Leu723 in the EGFR)	L755P was reported in lung cancer [95, 97, 107]. L755S was reported in gastric and breast cancers [109].	A model of ErbB2, based on the EGFR, predicted that Leu755 is located at the loop preceding the αC-helix. Leu755 is conserved in ErbBs from vertebrates, except for ErbB3, which displays isoleucine in this position.	The corresponding position to ErbB2's L755 in the EGFR is L723, which was also detected in cancer cells. (See above for analysis of this position.)
D769H (Asp737 in the EGFR)	D769H was reported in lung and gastric cancers [108, 109].	A model of ErbB2, based on the EGFR, predicted that Asp769 is located at the αC-helix, on the interface of the asymmetric dimer. The residue faces toward the solvent, not toward the hydrophobic packing with the activation loop. Asp769 is conserved in ErbBs from vertebrates.	D769H is predicted to affect formation of the active dimer.
V773A (Val741 in the EGFR)	V773A was reported in carcinoma of the head and neck [128].	A model of ErbB2, based on the EGFR, predicted that Val773 is located in the αC-helix, facing toward the protein. It is buried in the active	Because of its location on a regulatory element and its unique pattern of substitution, we predict that Val773 is functionally important and that its

		conformation, contacting a conserved motif at the N-terminus of the activation loop. Val773 is conserved as a valine residue in the EGFR and ErbB2 and as an isoleucine residue in ErbB4. In ErbB3 the corresponding position is mostly occupied by alanine (or serine in two fish orthologs).	substitution could be damaging. However, the nature of the effect of the substitution is not yet known.
G776S (Ser744 in the EGFR)	G776S was reported in gastric tumors [97].	A model of ErbB2, based on the EGFR, predicted that Gly776 is located at the C-terminus of the αC-helix.	The position corresponding to ErbB2's Gly766 in the EGFR is Ser744, which was also found to be mutated in cancer cells and was shown to increase the basal activity of the receptor. (See Table1 for analysis of this position.)
V777X X= [L, M] (Val745 in the EGFR)	V777L was reported in lung, colorectal and gastric cancers [107, 109]. V777M was reported in colorectal cancers [109].	A model of ErbB2, based on the EGFR, predicted that Val777 is located at the end of the αC-helix.	The position corresponding to ErbB2's Val777 in the EGFR is Val745, which was also detected in cancer cells. (See above for analysis of this position.)
Q799P (Gln767 in the EGFR)	Q799P was reported in gastric cancer [109].	A model of ErbB2, based on the EGFR, predicted that Gln799 is located on the interface of the symmetric putatively inactive dimer, contacts the C-terminal domain, and participates in the polar network of interactions across the dimer. Gln799 is conserved in ErbBs as a glutamine residue.	Q799P is predicted to lead to destabilization of the inactive dimeric state.

V842I (Val810 in the EGFR)	V842I was reported in colorectal cancer [109].	A model of ErbB2, based on the EGFR, predicted that the location of Val842 is close to the conserved catalytic loop.	The position corresponding to ErbB2's Val842 in the EGFR is Val810, which was also detected in cancer cells. (See above for the analysis of this position.)
N857S (Gln825 in the EGFR).	N857S was reported in an ovarian tumor [97].	A model of ErbB2, based on the EGFR, predicted that Asn857 is located on a loop on the back side of the catalytic site and is exposed to the solvent. Asn857 is not conserved in ErbBs.	Based on the evolutionary conservation analysis and its structural location, N857S is predicted to be a neutral mutation.
L869Q (Leu837 in the EGFR)	L869Q was reported in gastric cancer [109].	A model of ErbB2, based on the EGFR, predicted that Leu869 is located on the activation loop.	The position corresponding to ErbB2's Leu869 in the EGFR is Leu837, which was also detected in cancer cells and was shown to increase the basal activity of the receptor. (See above for the analysis.)
R896C (His864 in the EGFR)	R896C Was reported in breast cancer [109].	A model of ErbB2, based on the EGFR, predicted that Arg896 is located on a loop and is exposed to the surface. This residue is not conserved in ErbBs, but is evolutionarily variable.	Based on the conservation analysis and its structural location, this mutation is predicted to be neutral in its effect.
E914K (Glu882 in the EGFR)	E914K was reported in glioblastoma [97].	A model of ErbB2, based on the EGFR, predicted that Glu914 is located on a helix in the C-lobe. It is relatively buried in a hydrophobic environment, yet participates in forming an H-bond with the nitrogen of W856. Glu914 is conserved in ErbBs.	Based on the evolutionary conservation analysis this mutation is predicted to be damaging; however, the nature of the effect is not yet known.

In-frame Insertions/Duplications in *Exon20* in ErbB2			
774insAYVM (742 in the EGFR)	Reported in lung cancer [97, 116].	The insertions are located on the C-terminus of the αC-helix and its following loop and are close to the interface of the symmetric dimer.	All of these insertion are predicted to increase the basal activity and enhance transforming potential, similar to the G776insYVMA ErbB2 mutant [91] and the D746insNPG EGFR mutant [99] tested in vitro. Nevertheless, the molecular effects of these insertions might vary. The loop following the αC-helix in ErbB2 is known to bind Hsp90, which plays a role in restraining the activity of ErbB2, probably due to interference with formation of the asymmetric active dimer in which ErbB2 contacts the second monomer va its N-lobe. The alterations might have an effect on the active dimeric state.On the other hand, the two insertions that are located following residue 778 in ErbB2 (746 in the EGFR) might participate in destabilization of the inactive dimeric state.
775insYVMA (743 in EGFR)	Reported in lung cancer [106].		
G776XinsC (744 in EGFR) X = [L, V]	G776VinsC was reported in lung cancer [95]. G776LinsC was reported in lung cancer [105, 116].		
776-779insYVMA (S744−N747 in EGFR)	Reported in lung cancer [105, 107].		
779-781ins VGS (747−749 in EGFR)	Reported in lung cancer [97].		
781−783insGSP (749−751 in EGFR)	Reported in lung cancer [105].		

ErbB4 Alterations			
Missense Mutations			
V721I (Ile691 in the EGFR)	Reported by [111].	In a model of ErbB4 based on the EGFR, Val721 is located on the interface of the symmetric, putatively inactive dimer, close to Asp1012 (Glu981 in the EGFR) from the C-terminal domain. Val721 is conserved as a hydrophobic residue in ErbBs from vertebrates. It is substituted for an aspartate residue in two oncogenic EGFRs from southern platyfish *(Xiphophorus maculates)* and spiketail platyfish *(Xiphophorus xiphidium)*.	This mutation is predicted to lead to destabilization of the inactive dimeric state, although this prediction is weakened by the mild nature of the mutations.
A773S (Ala743 in the EGFR)	Reported by [111].	In a model of ErbB4 based on the EGFR, Ala773 is located on the αC-helix, on the interface of the asymmetric active dimer. It faces outward, contacting Ile949 (Ile917 in the EGFR) from the second monomer. Ala773 is conserved in all ErbBs except for ErbB3 orthologs that contain a glycine residue.	It is predicted to affect the formation of the active dimeric state. The exceptional residue occupying this position in ErbB3 might be related to the fact that ErbB3 probably does not form asymmetric heterodimers via its N-lobe (thereby acting as the activated monomer). Therefore, there is no evolutionary constraint in this position within ErbB3 orthologs, as opposed to the catalytically active ErbBs. This further indicates that the mutation in this residue in ErbB4 might have an effect on functionality.

R782Q (Arg752 in the EGFR)	Reported by [111].	In a model of ErbB4 based on the EGFR, Arg782 is located on the interface of the symmetric, putatively inactive dimer.	The position corresponding to ErbB4's Arg782 in the EGFR is Arg752, which was also detected in cancer cells. (See above for analysis of this position.)
E810K (Glu780 in the EGFR)	Reported by [111].	In a model of ErbB4 based on the EGFR, Glu810 is located on the surface of the protein, where it is exposed to the solvent. Glu810 is conserved in ErbBs from vertebrates, except for ErbB3, which displays a glutamine residue in this position. This position is also substituted for glutamine in oncogenic EGFRs from southern platyfish (*X. maculatus*) and spiketail platyfish (*X. xiphidium*).	Because of the unique conservation pattern of this position in the ErbBs, we predict that the mutation is damaging and will lead to enhanced activation.
P854Q (Pro824 in the EGFR)	Reported by [111].	In a model of ErbB4 based on the EGFR, Pro854 is located on the surface of the kinase domain; close to the interface with the C-terminal domain. The proline residue causes a turn in the structure between two β-strands. Pro854 is conserved in all ErbBs from vertebrates, except for ErbB3 from fish, which contain an aspartate residue in this position.	Because of the unique conservation pattern of this position in the ErbBs and its structural location, we predict that the mutation is damaging and will lead to enhanced activation.
D861Y (Asp831 in the EGFR)	Reported by [111].	Asp861 is part of the conserved DFG motif [127] at the N-terminus of the activation loop and is crucial for catalysis. Asp861 is conserved in	Because of the pattern of conservation and the structural location of this residue, the mutation is predicted to be damaging. However, the nature of the effect is not yet known.

		ErbBs and in tyrosine kinases in general.	
E872K (Glu842 in the EGFR)	Reported by [111].	In a model of ErbB4 based on the EGFR, Glu872 is located on the activation loop. Glu872 is conserved as a glutamate or aspartate residue in ErbBs from vertebrates.	Because of the pattern of conservation and the structural location of this residue, we predict that the mutation is damaging. However, the nature of the effect is not yet known.
T926M (Ala896 in the EGFR)	Reported by [111].	In a model of ErbB4 based on the EGFR, Thr926 is located on the surface of the protein, and is exposed to the solvent. Thr926 is not conserved in ErbBs.	Based on the conservation analysis and the structural location of this residue, the mutation is predicted to be neutral.
In-Frame Insertions/Duplications in *Exon20* in ErbB4			
G802insGGC (Gly772 in the EGFR)	Reported by [111].	In a model of ErbB4 based on the EGFR, Gly802 is located on a strand preceding the first helix in the C-lobe, in contact with the ATP analog [38]. Gly802 is conserved in ErbBs and is also largely conserved in tyrosine kinases in general.	Gly802 might be important for ligand binding and for catalysis. The insertion is predicted to be damaging, but the nature of the effect is not yet known.

Table 2: Alterations in EGFR are classified into three groups: missense mutations, *exon19* deletions, and *exon20* insertions. Alterations in ErbB2 and ErbB4 include missense mutations and *exon20* insertions. For alterations in EGFR, positions of the mutations in the mature EGFR are indicated (numbering in pre-mature EGFR is in parenthesis). For mutations in ErbB2 and ErbB4, the corresponding residue in the mature EGFR is indicated in parenthesis.

Chapter 2: Model Structure of the Na^+/H^+ Exchanger 1: Functional and Clinical Implications

This chapter is based on a published manuscript (Landau et al. (2007) *JBC* 282(52) 37854-63)

Abstract

Eukaryotic Na^+/H^+ exchangers are transmembrane proteins that are vital for cellular homeostasis and play key roles in pathological conditions such as cancer and heart diseases. Using the crystal structure of the Na^+/H^+ antiporter from *Escherichia coli* (EcNhaA) as a template, I predicted the 3-dimensional structure of the human Na^+/H^+ exchanger 1 (NHE1). Modeling was particularly challenging because of the extremely low sequence identity between these proteins, but the model-structure is supported by evolutionary conservation analysis and empirical data. It also revealed the location of the binding site of NHE inhibitors; which my colleagues and I validated by conducting mutagenesis studies with EcNhaA and its specific inhibitor 2-aminoperimidine. The model structure features a cluster of titratable residues that are evolutionarily conserved and are located in a conserved region in the center of the membrane; I suggest that they are involved in the cation binding and translocation. I also suggest a hypothetical alternating-access mechanism that involves conformational changes.

1. Introduction

Sodium/hydrogen transporters are ubiquitous transmembrane (TM) proteins that transport Na^+ and H^+ ions across the membrane, and are therefore imperative for vital cellular processes such as regulation of cellular pH, cell volume, and ion composition [129]. The mammalian Na^+/H^+ exchanger (NHE) family of transporters includes nine isoforms (NHE1 through NHE9), of which NHE1 is the most widely expressed. Following allosteric activation by intracellular acidification, NHE1 exchanges extracellular Na^+ for intracellular H^+ with Na^+:H^+ stoichiometry of 1:1 [130]. NHE1 is inhibited by amiloride and its derivatives and by benzoyl guanidium compounds such as cariporide [129]. Structurally, NHE1 is predicted to include two distinct domains: a TM N-terminal region of ~500 amino acids that is involved in ion translocation and drug recognition, and a cytoplasmic regulatory C-terminal domain of nearly 300 residues [131, 132]. The cytoplasmic domain includes the H^+ sensor and also serves to mediate regulation by other molecules or ions.

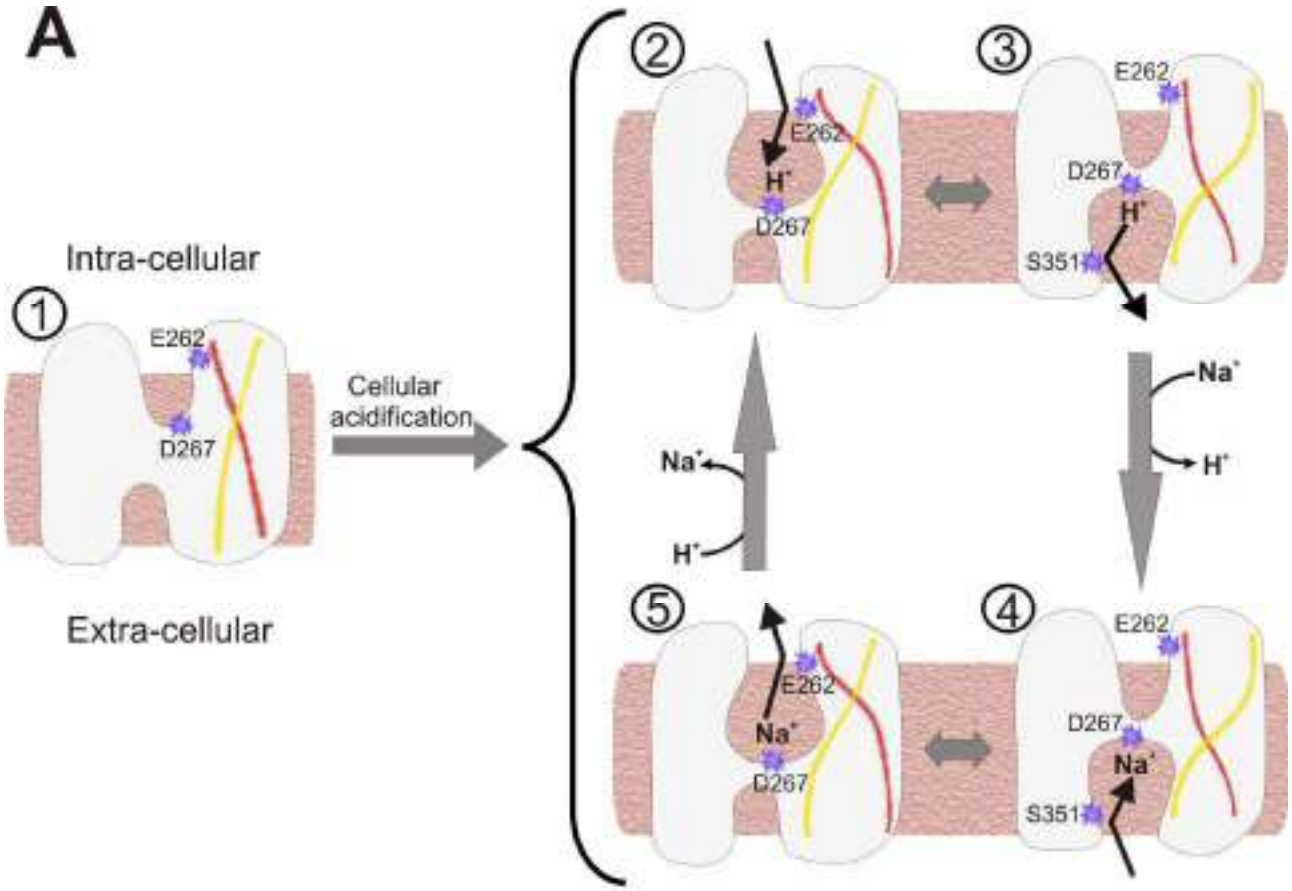
NHE1 is associated with many pathological conditions that include cancer as well as heart, vascular, gastric, and kidney diseases [129, 130]. For example, the activity of NHE1 is primarily involved in the damage inflicted on the human myocardium during and following a myocardial infarction, and accordingly, NHE1 inhibitors were shown to be beneficial during ischemia and reperfusion [129]. In addition, NHE1 plays a role in tumor growth by reversing the pH gradient in malignant cells, a phenomenon known as ‘malignant acidosis’, which is a key step in oncogenic transformation [129]. Therefore, NHE1 inhibitors can potentially serve as anti-cancer drugs [129].

NhaA, the main Na^+/H^+ antiporter in *Escherichia coli* (EcNhaA), is indispensable for bacterial growth in alkaline pH (in the presence of Na^+) and for adaptation to high salinity [133]. EcNhaA is an electrogenic antiporter extracting one Na^+ ion from the cell in return for inward current of two protons following cellular alkalization [133]. The function of EcNhaA is specifically inhibited by 2-aminoperimidine (AP), a guanidine-containing naphthalene derivative with some similarity to the NHE1 inhibitor amiloride [134]. The 3-dimensional (3D) structure of EcNhaA was recently determined, and found to comprise 12 TM segments [23].

The bacterial EcNhaA and eukaryotic Na^+/H^+ exchangers play similar roles in controlling pH and electrolyte homeostasis, and have been suggested to share a common ancestor and a similar structural fold [129, 135]. Thus, my working hypothesis was that

EcNhaA can be utilized as a template to predict the structure of the TM domain of NHE1. However, the proteins share very low sequence identity of about 10%, and it is not a simple matter to align their sequences and to predict the structure of NHE1 based on that of EcNhaA [136]. In this study, by using a fold-recognition approach, I obtained a 3D model of NHE1. Notably, the membrane topology of this model structure differs from the one that was suggested on the basis of hydrophobicity scales and cysteine accessibility analysis [137]. Reasons for the differences are discussed below.

My model of NHE1, which is supported by both phylogenetic and empirical data, incorporates the binding pocket of clinically important NHE inhibitors. This allowed me to locate the binding site of the AP inhibitor within the EcNhaA structure by site-directed mutagenesis. Finally, the integration of empirical data with the new structural model allowed me to suggest an alternating-access mechanism of the Na^+/H^+ exchange in molecular detail (Fig. 1A).


A
Intra-cellular
Extra-cellular
Cellular acidification
E262
D267
H+
Na+
S351

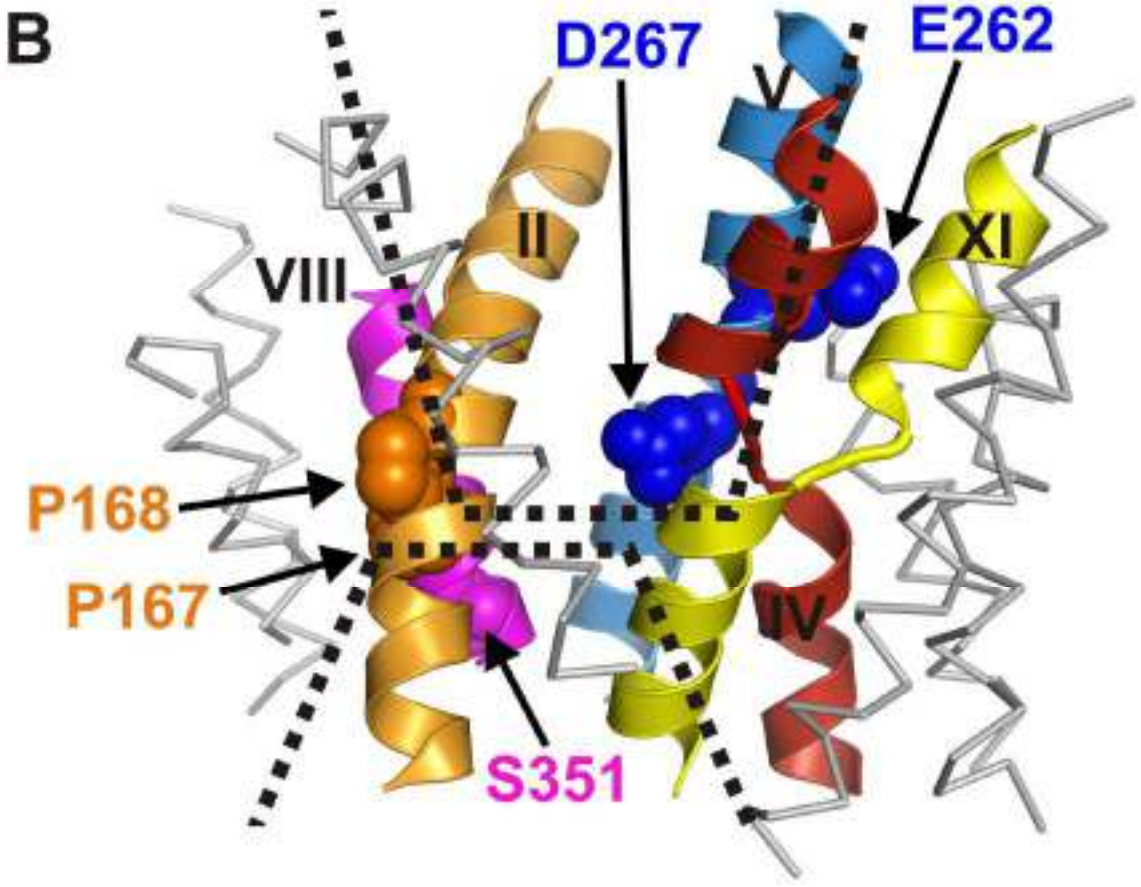

B
D267
E262
V
II
VIII
XI
P168
P167
IV
S351

Fig. 1. A Suggested Na^+/H^+ Exchange Mechanism of NHE1

(**A**) State 1 represents an inactive conformation, and the exchange cycle (states 2-5) illustrates putative conformational changes in the TM domain that follow activation by cellular acidification. The cycle involves dynamic equilibrium between conformations 2 and 5, in which the cation-binding site is accessible to the cytoplasm, and conformations 3 and 4, in which it is accessible to the extracellular matrix. The changes are mediated by the TM4−TM11 assembly and may also involve rotation of TM8 and exposure of Ser351 to the extracellular funnel (indicated in states 3 and 4). The cycle allows the transport of cations across the membrane via an alternating-access mechanism. In state 2, low pH promotes the entrance of a proton to the cytoplasmic funnel, probably attracted by the acidic Glu262, and the protonation of Asp267. The low pH also induces conformational changes, leading to the transfer of the proton from the cytoplasmic funnel to the extracellular funnel (state 3). In accordance with the chemical gradient of both cations, the proton is exchanged for sodium ion in the extracellular matrix, perhaps via Ser351 (state 4). Finally, movements to the alternative conformation (state 5), allows the replacement of sodium by a proton at the cytoplasmic side (state 2), again in accordance with their chemical gradients. The continuance of the cycle is controlled by cellular pH.

(B) The model structure of the TM domain of NHE1 in the inactive conformation of state 1 viewed from the membrane. The intracellular side is facing upward. TM segments that are important for function are represented by the colored ribbons. Other segments are represented by a gray trace. TM1 was omitted for clarity. Residues involved in the cation transport path are represented by space-filled atoms. The funnels laying the transport path are indicated by dashed lines.

---

2. Results

2.1. EcNhaA and Eukaryotic Na^+/H^+ Exchangers Share a Similar Fold

Using the sequence of NHE1 as a target, I detected EcNhaA as the closest homologue according to the fold-recognition FFAS03 server [138]. This finding strengthened my working hypothesis that the TM domains of the two exchangers share a similar fold.

2.2 Predicting the Topology of NHE1

2.2.1 Use of Multiple Approaches to Align the TM Domains of NHE1 and EcNhaA

The sequence identity between EcNhaA and human NHE1 is only ~10%, and I was unable to align their sequences using standard methods (data not shown). I therefore used several state-of-the-art approaches to construct alignments, and integrated the results. First, I extracted the pairwise alignment between NHE1 and EcNhaA, which displays 12.4% sequence identity, from a multiple-sequence alignment of a clan of transporters from the Pfam database [139]. Two additional pairwise alignments were calculated using

the FFAS03 [138] and HMAP [140] servers, which display 9.2% and 10.4% sequence identity, respectively.

2.2.2. TM Helix Assignment

I used each of the above alignments to assign the boundaries of 12 TM segments (TM1 through TM12) of NHE1, based on corresponding segments of the crystal structure of EcNhaA. Fig. 2 exemplifies the significant similarity between most of the TM segments predicted by the Pfam, FFAS03, and HMAP alignments. Using the iterative process described below, I predicted the final membrane topology (highlighted in yellow in Fig. 2 and illustrated in Fig. 3A).

Initially, the three different alignments were manually adjusted to reduce gaps in the TM helices of EcNhaA, and used to build 3D models of NHE1. The main dissimilarity between the different alignment methods appeared to be in the prediction of the TM6 and TM7 segments. The model structures provided additional information that was used to favor a specific assignment and improve it further; model structures that were favored were those with least polar residues facing the lipid bilayer. Such considerations favored adaptation of the Pfam assignment of TM6; they were not helpful, however, in assigning TM7, for which I therefore used information from a multiple-sequence alignment of homologous eukaryotic Na^+/H^+ exchangers. Because TM helices are expected not to include insertions and deletions of amino acids [136], I favored the assignment of TM7 to gap free region, as predicted by the FFAS03 and HMAP alignments but not by Pfam. Similar reasoning led me to reject the assignment of the first TM segment to residues 103−127—although that was the assignment predicted by all three methods (Fig. 2)—because this segment is highly variable and includes many insertions and deletions. In contrast, the next segment (residues 129−150), which was predicted by hydrophobicity analysis [137] to be a TM segment, is devoid of gaps. Interestingly, the conservation pattern in this region is compatible with the periodicity of a helix, i.e., a conserved residue appears at every fourth position, resulting in a conserved helical face (Fig. 3A). Accordingly, this was the region to which I assigned TM1.

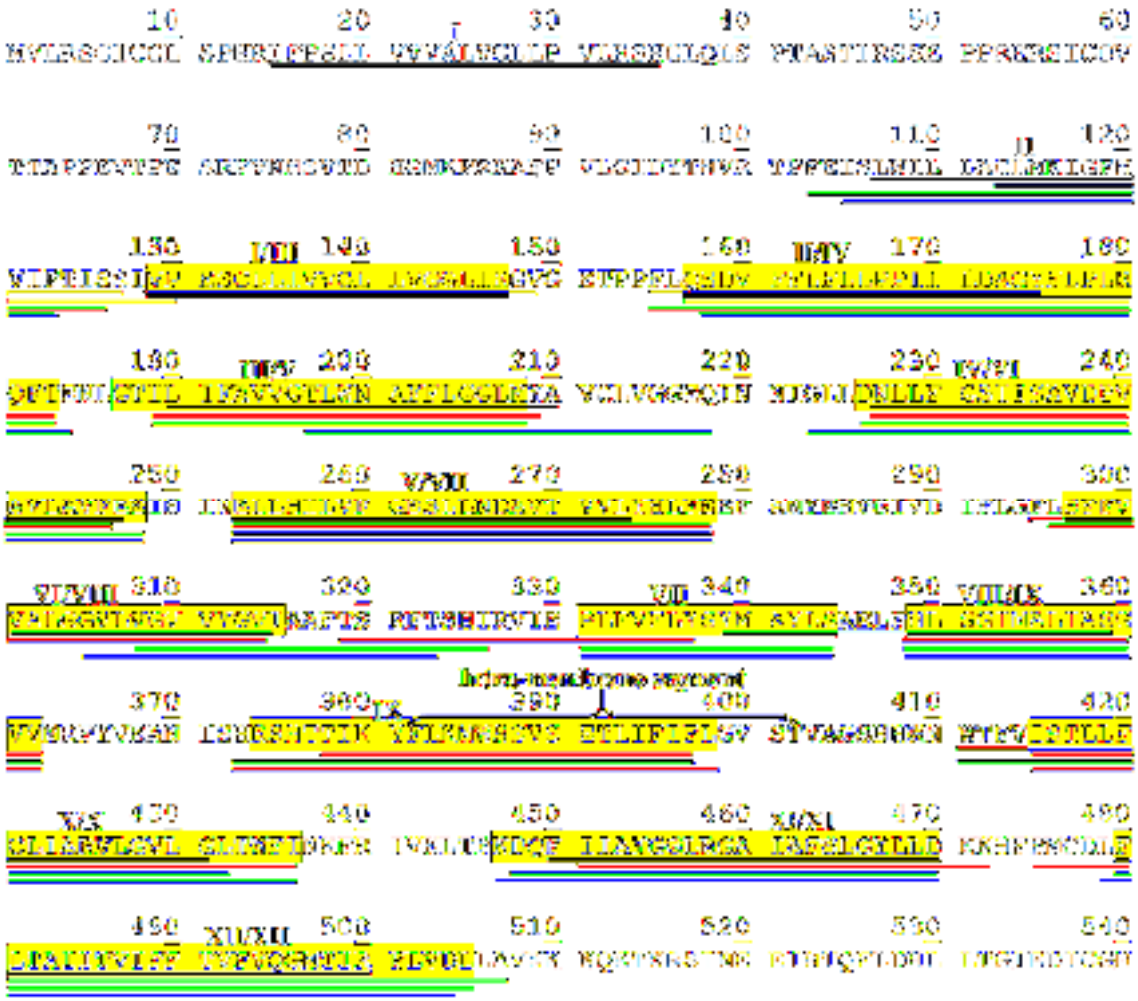


Fig. 2: The TM Segments in the NHE1 Sequence

The segments in NHE1 that correspond to the TM helices in EcNhaA, as predicted by the different methods discussed in the main text, are underlined on the sequence of NHE1 (residues 1-540) as follows: Pfam's prediction in red, FFAS03 in green and HMAP in blue. The boundaries of the TM segments as previously predicted [137], as well as their numbering, are indicated by black lines and Roman numerals, respectively. The segment predicted by Wakabayashi and co-workers [137] to be intra-membranal is also indicated. The final helix assignment proposed here is highlight in yellow and the numbering of the TM helices is indicated by the orange Roman numerals. The overall consensus between the methods is evident. The reasons for the selection of the location of TM1 are discussed in the main text.

2.3. Building the 3D model of NHE1

The above helix assignment of NHE1 was used to refine the pairwise alignment between NHE1 and EcNhaA in the TM regions. The final pairwise alignment displays 10.6% sequence identity (The alignment is available at my website http://bioinfo.tau.ac.il/~meytal/index_files/NHE/EcNhaA_NHE1_alignment.doc). A 3D model of NHE1 was subsequently constructed on the basis of this alignment and the EcNhaA template.

2.4. Assessment of the 3D Model

2.4.1. The 3D Model of NHE1 is Compatible with Evolutionary Conservation Analyses of Na^+/H^+ Exchangers

In helical proteins, evolutionarily conserved amino acids are typically located in strategic regions at the interfaces between the TM segments, whereas variable residues face the membrane lipids. The extra-membranal loops are also enriched in variable amino acids [21, 27, 28]. Accordingly, analyses of evolutionary conservation have been used to predict the structures of membrane proteins [24-26, 141]. They have also been exploited to validate model structures [21], as in the present study.

I projected the conservation scores calculated on the basis of the alignment of 94 sequences comprising the bacterial NhaA Na^+/H^+ antiporter family on the crystal structure of EcNhaA [23] (Figs. 3Ba and 3Bb). As expected, the most highly conserved residues are at the inter-helix interfaces within the TM region, while the most variable residues are located in the periphery; where they face the lipid membrane and populate the extra-membranal loops. Reassuringly, a very similar pattern was observed for my model structure of NHE1 (Figs. 3Bc and 3Bd). The results, obtained using an alignment of 305 Na^+/H^+ exchangers related to NHE1, strongly support my NHE1 model-structure.

Interestingly, a cluster of titratable residues (Figs. 4A and 4B), all evolutionarily conserved (Figs. 4C and 4D), is located within the conserved core in the center of the membrane in each of the structures. Titratable residues are very rare in the membrane, presumably because of the large desolvation free energy associated with their transfer from the aqueous phase into the membrane [142, 143]. Their presence in the membrane is often associated with function [144]. These titratable residues were indeed shown to be essential for the activity of both transporters [151-153], and arguably are involved in conformational changes and cation translocation (see Discussion and Fig. 1).

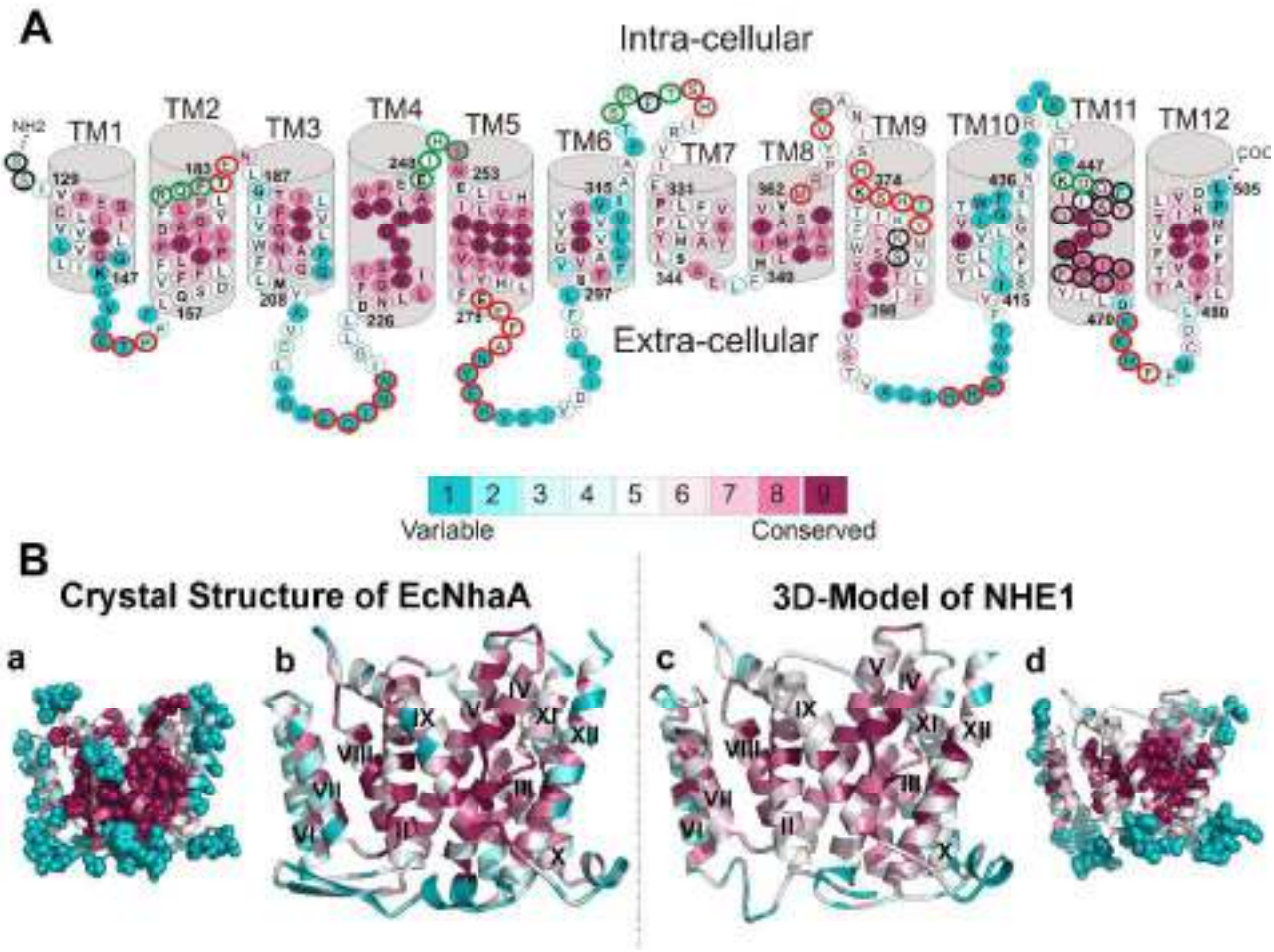


Fig. 3: Evolutionary Conservation Profiles of EcNhaA and NHE1

(**A**) The novel membrane-topology of NHE1 (residues 126-505) that I suggest here. The residues are colored according to their conservation-grades using the color-coding bar, with turquoise-through-maroon indicating variable-through-conserved. The start and end residue of each of the TM segments is marked in bold font and numbered. Residues that are located on the same helical face are situated on the same column (every fourth position). It is noteworthy that TM4 and TM11 unwind to form extended peptides within the helices. The results of the substituted cysteine accessibility analysis [137] are projected on the topology as follows: residues that are accessible to the intra- or extra-cellular medium are marked with green and red circles, respectively. The thick black circles mark residues that are completely inaccessible.

(**B**) The evolutionary conservation profiles of EcNhaA (left) and NHE1 (right) are projected on the crystal structure and 3D-model, respectively. The intracellular side is facing upward. The amino-acids are colored by their conservation-grades using the color-coding bar, as in (A). TM1 was omitted from the picture for clarity. **(b & c)** Ribbon models of EcNhaA and NHE1 viewed from the membrane. The TM segments are numbered. **(a & d)** The most variable (score 1) and conserved (score 9) residues are displayed by space-filled atoms. The compatibility of the NHE1 model structure with the phylogenetic profile is evident: The protein core is conserved while the periphery is variable, as with EcNhaA.

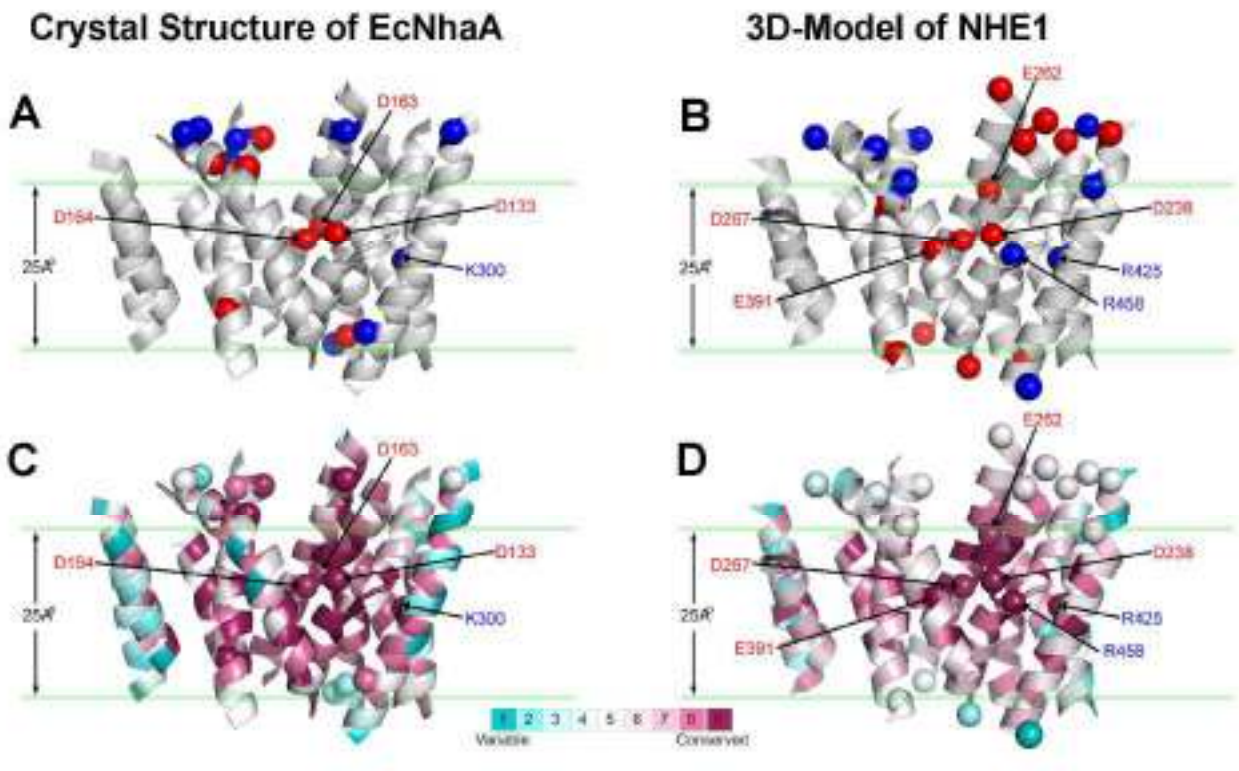


Fig. 4: Titratable Residues in the NHE1 and EcNhaA Transporters

A side-view of the crystal structure of EcNhaA [23] (**A & C**) and my model structure of NHE1 (**B & D**), which are displayed in a ribbon representation with the intracellular region in the upward direction. TM1 and the extra-membranal loops were omitted for clarity. The horizontal green lines mark the approximate boundaries of the hydrocarbon region of the membrane. In panel **A**−**B**, the transporters are colored gray, and the locations of the Cα atoms of the titratable residues are depicted as spheres. The red spheres correspond to aspartate and glutamate residues, and the blue to arginines and lysines. In panels **C**−**D** the amino-acids are colored by their conservation-grades using the color-coding bar, with turquoise-through-maroon indicating variable-through-conserved. Again, the locations of the Cα atoms of the titratable residues are depicted by spheres. It is evident that a central cluster of titratable residues is located in the conserved protein core, suggesting that it plays important functional roles in the transporters.

2.4.2. The NHE1 Model Structure is Consistent with the Positive-Inside Rule

Gunnar von Heijne and his co-workers showed that the topology of the vast majority of TM proteins is such that amino acid positions at the intracellular ends are enriched in the positively charged residues, lysine and arginine, relative to the extracellular side [143, 145]. This observation, termed the positive-inside rule, can be used to predict and evaluate the topology of membrane proteins. Analysis of the NHE1 3D model (incorporating residues 126−505) revealed 12 lysine/arginine residues on the cytoplasmic

side and only 3 lysine residues on the extracellular side (Fig. 3A). For reference, EcNhaA includes 16 lysine/arginine residues on the cytoplasmic side and 5 on the extracellular side [23].

2.4.3. The NHE1 Model Structure is Consistent with Mutagenesis Studies

Classical genetic and biochemical experiments and site–directed mutagenesis studies of eukaryotic Na^+/H^+ exchangers (listed in Tables 1 & 2) have yielded abundant data, which, for simplicity, I divided into two main groups: residues that are essential for function vs. those that are unessential. Residues were considered essential if their replacement resulted in loss or change of function (e.g., ion-translocation and pH-regulation), or if they were shown to be involved in binding of inhibitors.

When these mutagenesis data are projected on the NHE1 model structure, it can be seen that most of the residues defined as essential for activity are located in the core of the TM domain (Fig. 5), which is consistent with their role in maintaining the architecture and function of the transporter. On the other hand, most of the unessential residues face the membrane or are located in the extra-membranal loops. One essential residue, Ser351, unexpectedly faces the membrane lipids, and its functional relevance will be discussed below. Residues that participate in pH regulation, and thus mediate cellular signals, are located both on a cytoplasmic loop and within the protein core.

Mutagenesis studies point to 14 residues whose replacement affects the sensitivity of NHE1 to its inhibitors (Table 1). Some of these mutations do not affect Na^+ affinity, implying that the inhibitor-binding site is physically distinct and suggesting that the inhibitors induce allosteric regulation [146]. I focused on residues whose replacement significantly alters sensitivity to NHE inhibitors (i.e., by more than 10-fold), and which are likely to be directly involved in the binding. Specifically, mutagenesis implies that the binding site incorporates residues Phe161, Phe162 and Leu163, all located in TM2, and a second region comprising Gly352 of TM8 and Glu346 on its preceding loop (Fig. 5; Table 1). These two regions are located close to each other in my model, and Leu163 (TM2) is in direct contact with Glu346 and Gly352 (Fig. 5). Moreover, this binding site is situated at the extracellular side of NHE1, in accordance with the location of the inhibitors [147]. All in all, my NHE1 model structure is in excellent agreement with the mutagenesis data.

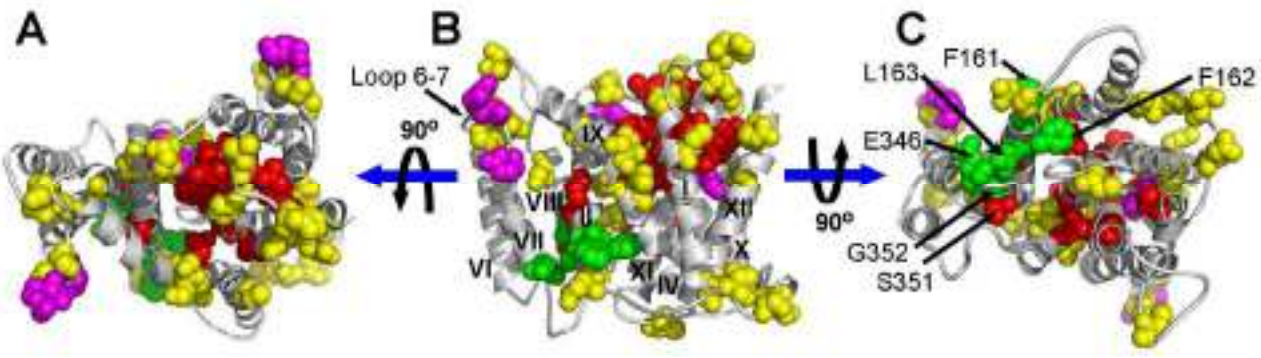


Fig. 5: Mutagenesis Studies in Eukaryotic Na^+/H^+ Exchangers

The 3D model-structure of NHE1 is displayed with gray ribbon and the residues that were mutated are presented using space-filled atoms using colors to represent the experimental outcome: Residues that were implicated in ion-translocation (P167, P168, S235, D238, P239, A244, L255, I257, V259, F260, G261, E262, N266, D267, T270, S351, E391, C421 and Y454) are colored red, residues that are involved in pH regulation (R180, R327, E330, R440, G455 and G456) in magenta, residues comprising the NHE-inhibitors binding-site (F161, F162, L163, E346 and G352) in green, and unessential residues (C133, Q157, P178, E184, C212, E248, H250, L254, H256, S263, V269, V271, F322, H325, S359, N370, S387, S388, S390, T392, S401, T402, S406, N410, K438, K443, C477, Q495 and R500) in yellow (for details see Tables 1 & 2). Residues involved in the binding of the NHE inhibitors, as well as Ser351 and loop 6-7 are labeled. These elements are discussed in the main text.

(**A**) A top view from the cytoplasmic side of the membrane. (**B**) A side view parallel to the membrane whereas the intracellular side is facing upward; the TM segments are numbered. **(C)** A view from the extracellular side.

2.5. NHE1 and EcNhaA Share a Similar Inhibitor-Binding Site

Mutations that alter the binding affinity of the NHE inhibitors were located in equivalent positions in a few eukaryotic NHE isoforms, implying that these isoforms share a common binding site (Tables 1 & 2). Thus, I assumed by extrapolation that the AP inhibitor of EcNhaA binds to an equivalent location on EcNhaA. Accordingly, my colleagues and I designed seven mutations in residues located in TM2 and TM8 of EcNhaA, and Katia Herz (under the supervision of Etana Padan) examined the sensitivity of their Na^+ or Li^+/H^+ activity to AP inhibition (Table 3; Fig. 6). The Na^+/H^+ antiport activity was measured in everted membrane vesicles isolated from EP432 transformed with the plasmids encoding the various mutations. EP432 lacks the chromosome-encoded antiporters (EcNhaA and EcNhaB) and expresses only the EcNhaA variants from a plasmid. Addition of the respiratory substrate, lactate, to these membrane vesicles (downward-facing arrow in Fig. 6) resulted in generation of ΔpH, as monitored by

quenching of the fluorescence of acridine orange, a fluorescent probe of ΔpH. Addition of either Na^+ or Li^+ to the reaction mixture (upward-facing arrow in Fig. 6) initiated the Na^+ or Li^+/H^+ antiport activity, as monitored by dequenching of the fluorescence. EP432 transformed with plasmid pAXH [148]) or the vector plasmid pBR322 served as positive and negative controls, respectively. To determine the effect of AP on the antiport activity, Katia Hertz added the inhibitor at various concentrations before adding lactate. The half-maximum inhibitory concentration (IC_{50}) of AP was determined as described [134].

Amino acid residues whose mutation changed the sensitivity to AP by at least threefold relative to the wild-type were considered to be involved in or affect the AP binding site. Specifically, mutations W62C, F72C, G76C and H225R exerted no effect on inhibition by AP; the IC_{50} value was very similar to that of the wild type (1.7 μM). N64C and F71C mutations increased the sensitivity of the Na^+ but not of the Li^+ antiport activity to AP inhibition; the IC_{50} values of AP for these mutants were 0.5 μM and 0.3 μM, respectively. In contrast, N64C and H225Q decreased the sensitivity of the Li^+ but not of the Na^+ antiport activity to AP inhibition; the IC_{50} values of AP for these mutants were 17 μM and 7.8 μM, compared to 2.2 μM for the wild type.

The above results support our conjectured location of binding of the AP inhibitor on EcNhaA. We cannot yet explain why the substitution of Cys for N64 and F71 increased the sensitivity of the Na^+ but not of the Li^+ antiport activity to AP whereas similar mutations in N64 and in H225 decreased the sensitivity of the Li^+ but not of the Na^+ antiport activity. We can only speculate that the binding sites of Li^+ and Na^+ differ in size, as suggested previously [149] and as predicted from the different sizes of these hydrated cations.

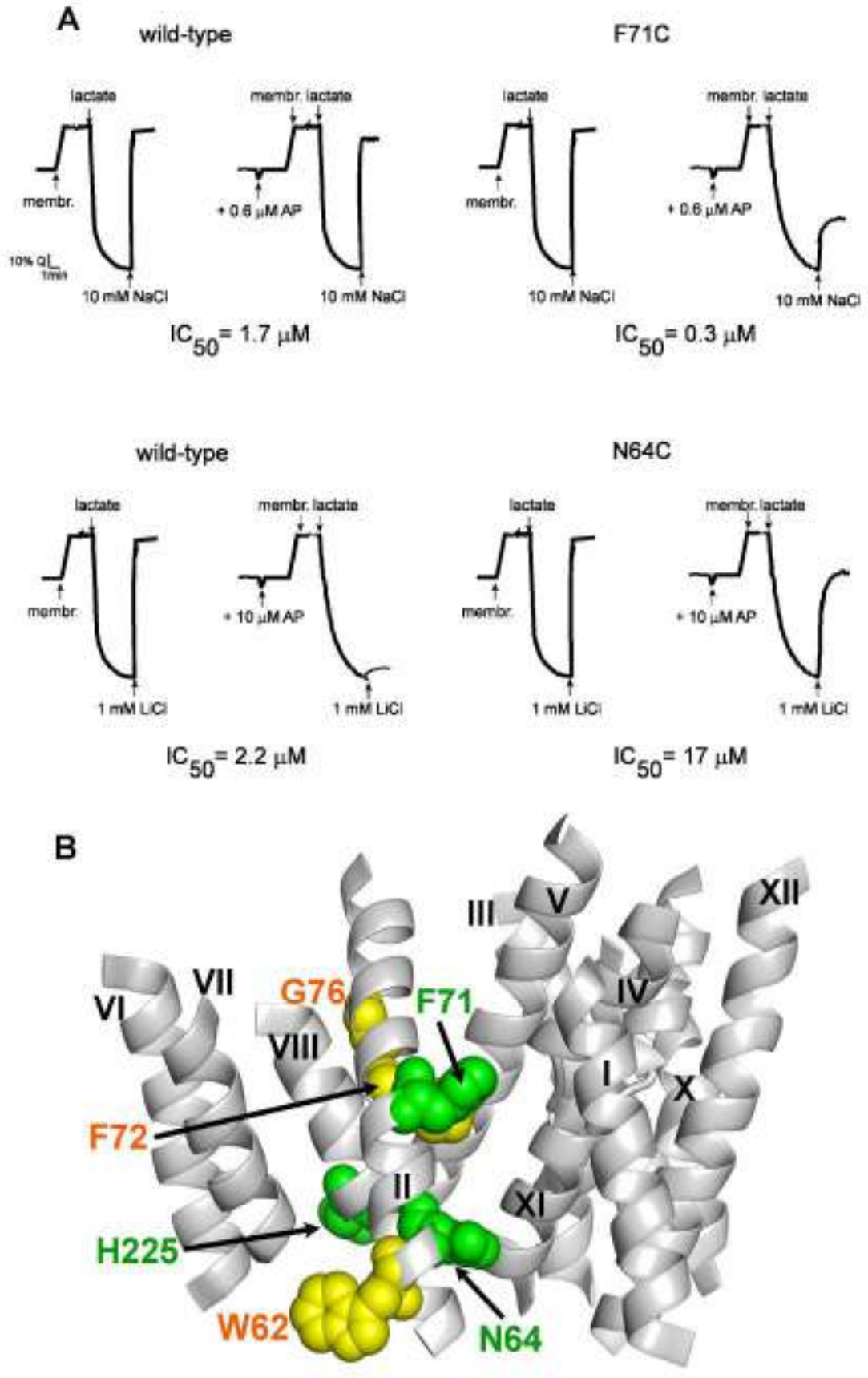

A
wild-type
F71C
lactate
membr.
membr. lactate
+ 0.6 µM AP
10% Q
1 min
10 mM NaCl
IC50= 1.7 µM
IC50= 0.3 µM
wild-type
N64C
+ 10 µM AP
1 mM LiCl
IC50= 2.2 µM
IC50= 17 µM
B
III
V
XII
VII
VI
G76
F71
IV
VIII
I
X
F72
II
XI
H225
W62
N64

Fig. 6: The Binding Site of the AP Inhibitor of EcNhaA

(A) The effect of AP inhibition on the antiporter activity of EcNhaA mutants compared to wild-type. The results obtained for the F71C and N64C mutations showing the most drastic AP effect, are displayed. Everted membrane vesicles were isolated from EP432 cells expressing wild-type NhaA or the indicated mutants, and the Na^+/H^+ or Li^+/H^+ antiporter activity was measured at pH 7.5. At the onset of the reaction, membranes were added first and then Tris D-lactate (2 mM) (↓), and the fluorescent quenching (*Q*) was recorded until a steady state level of ΔpH (100% quenching) was reached. NaCl or LiCl, at the indicated concentrations, was then added (↑), and the new steady state of fluorescence obtained (dequenching) after each addition was monitored. Where indicated, AP, at the indicated concentrations, was added (↑), to the reaction mixture following the addition of the membranes. The experiments were repeated at least three times with practically identical results. Calculated IC_{50} is shown for each experiment.

(B) The crystal structure of EcNhaA [23] is displayed in a gray ribbon representation. Space-filled atoms represent residues that were examined for their involvement in AP binding (Table 1). Residues that play a role in mediating AP inhibition (Asn64, Phe71 and His225) are colored green, while naïve residues (Trp62, Phe72 and Gly76) are colored yellow.

2.6. Comparison between Novel and Previously Suggested NHE1 Topologies

My model, which was derived from sequence alignments with a functional homologue, presents a novel topology. A previously suggested topology, which was based on hydrophobicity scales [137] (Fig. 2), was assessed by substituting cysteines for 83 of NHE1's residues and determining the accessibility of these substituted cysteines to cysteine-directed reagents from outside and inside the cell [137]. That analysis (Fig. 3A) yielded conflicting results in two regions, where accessibility to both sides of the membrane was apparent in adjacent residues. The authors suggested that such regions could be inserted into the membrane and might play a role in ion translocation. I believe that this is indeed the case in one of these regions, which in my model is located at the end of TM2 (see Discussion). The second region, located in the loop between TM6 and TM7, might play a role in pH regulation (Table 1 and Fig. 5).

Both of the predicted topologies assigned the location of the N- and C-termini of NHE1 inside the cytoplasm, in accordance with the topology of EcNhaA and experimental evidence [131, 132]. However, whereas the first TM segment in my model begins at residue Val129, Wakabayashi and co-workers predicted two additional segments at the preceding N-terminal end (Fig. 2). The N-terminal segment of NHE1 was suggested to serve as a signal sequence [150], and consistently with that suggestion my

evolutionary conservation analysis disclosed that this region (the first ~110 residues) is highly variable among Na^+/H^+ exchangers, and in some of them is even missing. Chymotryptic cleavage of the N-terminal region (residues ~1−150), following expression of NHE1, indeed had little effect on transport activity [132].

As shown in Fig. 2, although the suggested topology [137] contained two extra TM segments at the N-terminus, the next six predicted TM helices overlap with my model. Similarly, within the two topologies the last three helices coincide (Fig. 2). The remaining three segments (TM7−TM9) differ between the topologies. TM7 and TM8 in my model are predicted to be short (14 residues) relative to the other segments (19−27 residues). In contrast, the ninth TM segment predicted by Wakabayashi and co-workers is of normal length, and encompasses roughly these two short TM segments. This assignment is presumably due to the constant size of the window that is used in common hydropathy plots.

The intracellular region that follows TM8 in my model unexpectedly displays residues that are accessible to external reagents, followed by inaccessible residues (TM9) and then again by residues accessible to external reagents (the extracellular loop between TM9 and TM10). Wakabayashi and co-workers resolved this inconsistency by assigning an intra-membrane span instead of a transmembranal one (Fig. 2). An intra-membrane span in this region was also suggested for the *Arabidopsis thaliana* Na^+/H^+ exchanger isoform 1 (AtNHX1) [150]. Alternatively, it is possible that the intracellular loop between TM8 and TM9 participates in ion translocation and is therefore accessible to external reagents. My model structure is consistent with the latter possibility.

3. Discussion

Technical difficulties in experimental determination of the membrane topology and 3D structure of NHE1 prompted me to use computational tools to predict its structure based on the crystal structure of the prokaryotic EcNhaA antiporter. This is not a simple undertaking because of the extremely low sequence identity between these two proteins, and necessitated manual integration of the results of various computational tools. The resulting NHE1 model structure is supported by evolutionary conservation analysis and empirical data, as elaborated in Results, suggesting that it represents a fair approximation of the real structure of this protein. In addition, with the help of my colleagues, I located the binding site for inhibitors in both NHE1 and EcNhaA. The finding that the NHE1 and

EcNhaA transporters, from human and bacteria, respectively, share a common binding site for inhibitors provides strong support for my contention that the 3D structure of the former can be based on the crystal structure of the latter, despite their low sequence similarity.

3.1. Functional Implications of the Model: Similarity to EcNhaA

The most notably conserved helices in both NHE1 and EcNhaA are TM2, TM4, TM5, TM8, and TM11, all located in the protein core (Fig. 3). I suggest below that the similarity in conservation patterns of the two proteins, as well as the equivalent locations of functionally important sites within them (Fig. 4), indicate that they share similar transport mechanisms.

3.1.1. The TM4 and TM11 Assembly Lays the Core of an Alternating-Access Mechanism

EcNhaA includes an assembly of the TM4 and TM11 segments, both unwinding to form extended peptides in the center of the helix, which cross each other in the middle of the membrane [23]. These irregular structures form dipoles that are stabilized by two titratable residues located on TM4 (Asp133) and TM10 (Lys300) [23] (Figs. 4A and 4C). Their positions are conserved in bacterial NhaA Na^+/H^+ antiporters as aspartate and lysine residues respectively, and were shown to be essential for activity [151-153]. In the NHE1 model structure the corresponding positions include, respectively, the essential Asp238 (Table 2) and Arg425, (Fig. 4B), which are highly conserved among NHE1-related Na^+/H^+ exchangers (Fig. 4D) as aspartate and arginine residues, respectively, and can also compensate for the helix dipoles. In addition, the region predicted to unwind within TM11 in NHE1 contains two essential and conserved glycine residues (Table 1). Glycines are not favored in helical structures, and their presence in this region of the model structure might facilitate unwinding of the TM11 helix. Another conserved and potentially charged residue, Arg458, is located on TM11 of NHE1 (Figs. 4B and 4D). Its mutations to cysteine abolishes expression of NHE1 [137], implying that it is structurally important.

EcNhaA displays two funnels that were suggested to lay the ion-transport path [23]. One funnel, open to the cytoplasm, is formed by the cytoplasmic parts of TM2, TM4, TM5, and TM9. The other, open to the periplasm, is formed by the periplasmic parts of TM2, TM8, and TM11. Both funnels are blocked in the middle of the membrane near the

TM4−TM11 assembly, and do not form a continuous pore [23]. Hunte and coworkers suggested that conformational changes following pH activation presumably allow two alternating conformations of the cation-binding site at the bottom of the funnels. They pointed out that the TM4−TM11 assembly might lay the core of the alternating-access mechanism, as the extended peptides in the middle of the membrane and their dipoles are eminently capable of subtle and rapid conformational changes in response to activation [23].

I suggest that the TM4−TM11 assembly plays a similar role in NHE1 because of its high conservation and the clustering of essential residues within these segments, both critical factors in maintaining the structure, ion-transport, cation-specificity and selectivity, and pH regulation of eukaryotic Na^+/H^+ exchangers (Tables 1 & 2).

3.1.2. TM2 Shapes the Path of Cation Transport

TM2 contributes to the formation of the two funnels that are suggested to lay the path for cation transport [23, 153] (Fig. 1B). The crystal structure of EcNhaA shows a bending of the TM2 helix that is probably important for its structural role. Interestingly, the TM2 segment in my NHE1 model structure contains two proline residues (Pro167 and Pro168; Fig. 1B), both of which are essential (Table 1), and an additional Pro178 located at the cytoplasmic end. Proline-rich segments correspond to irregular helical structures [142], and the assignment of this helix nicely fits its structural features and supports my model structure. In this respect, it is noteworthy that a peptide that corresponds to TM2 in my model (previously known as TM4), and which was studied using high-resolution NMR spectroscopy, displayed irregular structural properties [154]. Overall, the location of TM2 in my model structure, in combination with the mutagenesis data showing that substitutions of residues located on this segment result in a nearly inactive protein in many cases (Tables 1 & 2), implying that this segment might lay the cation-transport path, as suggested [154].

3.1.3. Residues in TM5 Serve as the Cation-Binding Site

TM5 is located spatially close to TM4 and TM11 and displays exceptionally high evolutionary conservation, mainly in residues facing the TM4−TM11 assembly, both in EcNhaA and NHE1 (Fig. 3). Extensive mutagenesis within TM5 in NHE1 demonstrated its importance for expression and targeting to the membrane [146], in accordance with its strategic location in the protein core in my model.

In EcNhaA, this helix includes two titratable residues, Asp163 and Asp164 (Fig. 4A and 4C). These residues, which are located in the middle of the membrane, are evolutionarily conserved in the bacterial NhaA Na^+/H^+ antiporters as aspartate residues, are essential, and are considered to be the cation-binding site [23, 152]. According to my model structure, the corresponding residues in NHE1 are Asn266 and Asp267, respectively. These residues are highly conserved as aspargine and aspartate residues, respectively, within the family of NHE1-related Na^+/H^+ exchangers. Asp267 is located at the bottom of the cytoplasmic funnel (Fig. 1B) and, by conjecture, is involved in cation binding; a negative charge at this position is indeed crucial for function [146], and even a mild substitution abolishes the activities of NHE1 and its yeast homologue sod2 (Tables 1 & 2).

3.1.4. Is TM5 Responsible for the Different Stoichiometries in EcNhaA and NHE1?

I postulate that replacement of the negatively charged Asp163 in EcNhaA by the neutral residue Asn266 from NHE1 is important for the observed difference in Na^+:H^+ stoichiometry between these two transporters (1:2 in EcNhaA and 1:1 in NHE1). According to my hypothesis, Asp164 in EcNhaA or Asp267 in NHE1 serves to alternately bind Na^+ or H^+ (Fig. 1). On the other hand, Asp163 in EcNhaA binds the second proton, while its equivalent in NHE1—Asn266—does not participate in cation binding. Nevertheless, because Asn266 is conserved, essential (Table 1), and facing adjacent helices, I suggest that it might be of structural importance. Further mutations in these residues in both EcNhaA and NHE1 are likely to shed some light on their role in determining the Na^+:H^+ stoichiometry of the transporters.

3.1.5. Titratable Residues, Unique to NHE1, Putatively Involved in Ion-translocation

Besides Asn266 and Asp267 discussed above, another acidic residue in TM5, namely Glu262, was also shown to be crucial for the function of NHE1 and its yeast homologue, sod2 (Tables 1 & 2). This residue is fully conserved as glutamate in NHE1-related Na^+/H^+ exchangers, and its proximity to the cytoplasm (Fig. 4B and 4D) suggests that it might attract protons following cellular acidification (Fig. 1A).

Glu391 (TM9), which faces the cytoplasmic funnel, spatially close to the TM4−TM11 assembly (Fig. 4B and 4D), might play a role in the ion-translocation pathway. This position is conserved in NHE1-related Na^+/H^+ exchangers, where it is largely occupied

by glutamate but also by either asparagine or glutamine. Substitution of glutamine for Glu391 in NHE1 significantly reduced activity but did not abolish it (Table 1), implying that this residue is important in NHE1 but is not the main binding site.

Overall, a cluster of three conserved acidic residues (Glu262, Asp267, and Glu391) is located within the core of NHE1 (Fig. 4B and 4D), and I suggest that it facilitates the binding and translocation of the cations in this transporter, whereas Asp267, located at the bottom of the cytoplasmic funnel, serves as the main cation-binding site (Fig. 1).

3.1.6. TM8 Plays a Role in NHE1 Activity

TM8 is highly conserved both among the bacterial NhaA Na^+/H^+ antiporters and among NHE1-related Na^+/H^+ exchangers (Fig. 3). Within this short helix, it is especially noteworthy that His225 in EcNhaA and its equivalent Ser351 in NHE1 (Fig. 5), despite their conservation and polarity, face the membrane. This observation is especially interesting in view of the demonstration by mutagenesis analyses that different substitutions in His225 shift activity to more acidic pH (H225R), or to more alkaline pH (H225D), or abolish the activity of EcNhaA completely (H225A) [133, 155]. Rotation of TM8 by approximately 180° would place the side-chains of His225 in EcNhaA, and of Ser351 in NHE1, in the external funnel. I therefore suggest that activation of these transporters involves rotation of TM8 around its axis, such that these residues can participate in cation transport (Fig. 1).

Ser351 in NHE1 and its adjacent neighbor Gly352, which is essential (Table 1), are both highly conserved in NHE1-related Na^+/H^+ exchangers, mostly as serine and glycine residues. Interestingly, my evolutionary conservation analyses showed substitutions of aspartate residues for both Ser351 and Gly352 in a fungi-specific clade of plasma membrane Na^+/H^+ exchangers. The aspartate pair was shown to be important for activity in several of these transporters (Table 2). The unique identity of these residues implies a specific trait that is attributed to fungi exchangers, e.g., in mediating cation transport, and supports the importance of TM8 for activity.

3.2. A Putative Exchange Mechanism in NHE1

The scheme in Fig. 1A summarizes the suggested alternating-access mechanism for Na^+/H^+ exchange in NHE, and Fig. 1B highlights the location of the main residues that are implicated in the exchange. Overall, the mechanism, which involves consecutive transformations between pairs of conformations that are at chemical equilibrium, is

driven by the concentration gradients of sodium or protons across the membrane. The cation-transport path is formed by two discontinuous funnels comprised of TM2, TM4, TM5, and TM9 at the cytoplasmic side and TM2, TM8 and TM11 at the extracellular side. Upon activation by intracellular acidification, a proton, possibly attracted by Glu262 (TM5), enters the cytoplasmic funnel and binds to Asp267 (TM5) located at the bottom of the funnel (state 2 in Fig. 1A). Conformational changes, induced by the TM4−TM11 assembly, which might include rotation of TM8 by approximately 180°, then shield the proton from the cytoplasm. Alternatively, an external path now opens to the extracellular matrix (state 3 in Fig. 1A), which is enriched with sodium. A sodium ion can now compete with the proton for binding to the extracellular site, possibly mediated by Ser351 (TM8) (state 4 in Fig. 1A). Binding of sodium favors the movement to the alternative conformation, which shields the sodium cation from the extracellular matrix and opens the path to the cytoplasm (state 5 in Fig. 1A). The sodium can then be released and replaced by a proton, again in accordance with the chemical gradient of these cations (state 2 in Fig. 1A), and the cycle continues.

4. Concluding remarks

My model structure of NHE1 is supported by phylogenetic and published empirical data available for NHE1 and other eukaryotic Na^+/H^+ exchangers, specifically pertaining to the protein core (TM2, TM4, TM5, TM8, and TM11). These central segments are evolutionarily conserved and include essential residues in the NHE1 and EcNhaA transporters. Moreover, both transporters display a cluster of titratable residues in the center of the conserved protein core (Fig. 4) that are essential (Tables 1 & 2) and are presumably involved in conformational changes and cation translocation. Thus, my colleagues and I are fairly confident of the correctness of my model structure of the core of the NHE1 transporter.

On the other hand, I note that the location of the peripheral TM segments TM1, TM3, TM6, TM7, TM9, TM10, and TM12 in the model structure might be approximate, and that the conformations of the extra-membranal loops are tentative. Additional structural data, e.g., from high-resolution cryo-EM and X-ray crystallography, are needed in order to further our knowledge of these regions. Nevertheless, the results of this study demonstrated that even a model structure, particularly when integrated with experimental data, can be used to propose testable hypotheses that will ultimately shed light on

function and regulatory mechanisms. They might also pave the way to structure-based drug design, yielding additional NHE1 inhibitors of clinical benefit.

5. Methods

5.1. Classification and Nomenclature of Na^+/H^+ transporters

5.1.1. NHE1 and EcNhaA Belong to Two Different Protein Families

NHE1 is a part of the CPA family

TCDB (http://www.tcdb.org/tcdb/) is a database that classifies transport proteins [156]. In this database, membrane proteins that transport monovalent cations in exchange for protons belong to the CPA (Cation:Proton Antiporter (TC#2.A.37)) superfamily [156]. They are roughly assigned to three families: CPA1 (TC#2.A.36), CPA2 (TC#2.A.37), and the Na^+-transporting carboxylic acid decarboxylase (NaT-DC) family (TC#3.B.1).

The CPA1 and CPA2 families contain exchangers from bacteria, archaea, and eukaryotes. Members of the CPA1 family mostly catalyze Na^+:H^+ exchange, although some might also transport Li^+ or even Ca^{+2} instead of Na^+ [156]. The CPA1 family includes the functionally characterized mammalian Na^+/H^+ exchanger (NHE) isoforms, also referred to as the solute carrier 9A (SLC9A) by the HUGO nomenclature [157] (http://www.gene.ucl.ac.uk/nomenclature), and the plant and yeast NHX transporters. Members of the CPA2 family catalyze both K^+/H^+ and Na^+/H^+ antiport, as well as K^+ uniport, and include the KefB/KefC K^+ efflux proteins of *E. coli* that are responsible for glutathione-gated K^+ efflux [156]. Members of the NaT-DC family are so far known to contain only bacterial proteins. They export Na^+ from the cell using the energy from decarboxylation of a carboxylic acid substrate.

EcNhaA is a part of the NhaA Na^+:H^+ antiporter family

EcNhaA (the main Na^+/H^+ antiporter in *Escherichia coli*) is classified in TCDB as a unique family called the NhaA Na^+:H^+ antiporter (NhaA) family (TC#2.A.33).

5.1.2. NHE1 and EcNhaA Belong to the Same Superfamily of Transporters

Pfam is a comprehensive database for protein domains and families. Protein families that have arisen from a single evolutionary origin are clustered in clans. The CPA1 and CPA2 families (that include NHE1), the K^+ transporter (Trk) family and the NhaA family (that include EcNhaA) are classified into one Pfam clan (the CPA/AT clan) along with proteins

that are involved in sodium transport and other proteins of unknown function [158]. Pfam divides this clan into 10 different families. The sodium/hydrogen exchanger family (Pfam accession no. PF00999) includes proteins from the CPA1 and CPA2 TCDB families. EcNhaA belongs to a different Pfam family, the Na^+/H^+ antiporter 1 family (Pfam accession no. PF06965), which includes only bacterial proteins.

5.2. Evolutionary Conservation Analysis of the NhaA Na^+/H^+ Antiporter Family

The Pfam [139] alignment for the NhaA Na^+/H^+ antiporter 1 family (Pfam accession no. PF06965), comprising 198 sequences, was taken as the initial set of homologous sequences. Redundant (>95% sequence identity) and fragmented sequences were discarded. This intermediate multiple sequence alignment (MSA) of 136 sequences was used to generate a Hidden Markov Model [159], which was subsequently utilized to collect homologous sequences from the UNIPROT database [118]. Full-length sequences of these proteins were aligned using the MUSCLE program [120]. The N- and C-terminal positions of the alignments were then cut to fit the sequence of EcNhaA (residues 1−388) (SwissProt entry: NHAA_ECOLI). Redundant (>95% sequence identity) and fragmented sequences, sequence variants, and mutants were discarded. Also discarded were proteins that were sequenced by the Whole-genome Shotgun (WGS) project and are therefore considered as preliminary data. The resultant MSA contained 94 sequences of bacterial proteins.

Based on the 94-sequence MSA, evolutionary conservation scores were calculated using a Bayesian method [121] and projected onto the 3D structure of EcNhaA (PDB entry 1ZCD; [23]) using the ConSurf web server (http://consurf.tau.ac.il/) [5].

5.3. Evolutionary Conservation Analysis of NHE1-Related Na^+/H^+ Exchangers

The ConSeq web server (http://conseq.bioinfo.tau.ac.il/) [10] was used to generate an initial alignment of Na^+/H^+ exchangers as follows. Human NHE1 (residues 105−505) (SwissProt entry: SL9A1_HUMAN) was used as a query sequence to collect homologous sequences from the UNIPROT database [118] using PSI-BLAST [119]. The resulting 216 sequences were aligned using MUSCLE [120] with default parameters. Redundant (>95% sequence identity) and fragmented sequences were discarded. The resultant intermediate MSA of 151 sequences was used to generate a Hidden Markov Model [159], which was subsequently utilized to collect remote homologous sequences from the UNIPROT database [118] that

were aligned using MUSCLE [120]. From the 516 hits found, redundant (>95% sequence identity) and fragmented sequences, sequence variants, and mutants were discarded along with sequences that included irregular characters or ones that were sequenced by WGS project. I then limited the boundaries of the MSA to positions corresponding to residues 126−505 of NHE1, which corresponds to my 3D model of NHE1 described in the main text. The final alignment of 305 proteins from all kingdoms was exclusive to Na^+/H^+ exchangers related to NHE1 and highly reliable to infer position-specific evolutionary information for this transporter.

Based on the 305-sequence MSA, evolutionary conservation scores were calculated using a Bayesian method [121], and, using the ConSurf web-server (http://consurf.tau.ac.il/) [5], were projected onto the 3D model of NHE1 described below.

5.4. Identifying the Fold of Eukaryotic Na^+/H^+ Exchangers

Neither the search performed with PSI-BLAST nor the Hidden Markov Model built for NHE1-related Na^+/H^+ exchangers could detect sequences from the prokaryotic NhaA Na^+/H^+ antiporter family, probably because of low sequence similarity. Therefore, to detect those protein structures most likely to possess a similar fold to that of eukaryotic Na^+/H^+ exchangers, I used profile-to-profile alignments implemented in the FFAS03 server [138]. The sequences of both human NHE1 (SwissProt entry: SL9A1_HUMAN) and the yeast Na^+/H^+ transporter sod2 (SwissProt entry: NAH_SCHPO) were used as queries. In both cases, EcNhaA was the closest homologue whose 3D structure is known.

5.5. Pairwise Sequence Alignment between Human NHE1 and EcNhaA

Generally, and in particular when the sequence homology is low, the use of MSAs to deduce a particular alignment between two sequences produces a more accurate result than the use of a simple pairwise alignment [22, 140, 160]. Moreover, it was previously shown that construction of alignments using sequence profiles and Hidden Markov Models [159] results in a more accurate match than using standard sequence-based alignments [160-163]. The Pfam database [139] provides state-of-the-art Hidden Markov Model-based alignments for protein domains that are predicted to arise from a single evolutionary origin, referred to as clans [158]. I therefore used the Pfam's MSA for the clan comprising both EcNhaA and NHE1, and extracted 107 sequences including the bacterial NhaA Na^+/H^+ antiporter 1 family (22 sequences including EcNhaA) and the CPA $Na^{+/}H^+$ exchanger family (85 sequences including NHE1). The alignment was also recalculated using MUSCLE [120] to provide an

additional alignment method, thereby substantiating the alignment's accuracy. Two pairwise sequence alignment between human NHE1 (SwissProt entry: SL9A1_HUMAN) and EcNhaA (SwissProt entry: NHAA_ECOLI) were directly deduced from the MSAs constructed by Pfam or MUSCLE, as described above. Gaps in corresponding positions were eliminated. Because both alignments yielded similar results, I made no further use of the MUSCLE alignment.

The FFAS03 server applies profile-to-profile alignments and a fold recognition algorithm to detect and align structurally determined proteins to a given sequence [138]. As mentioned above, this method detected EcNhaA as the closest structurally determined homologue to NHE1. I therefore used the resulting sequence alignment between these proteins as an additional pairwise alignment.

The HMAP server [140] offers an additional state-of-the-art approach, which I used to calculate the alignment between NHE1 and EcNhaA. HMAP stands for 'hybrid multidimensional alignment of profiles'. This approach combines information on sequence and structure (including secondary and tertiary structures) in order to align remote homologous sequences as accurately as possible [140, 161, 162]. Human NHE1 (SwissProt entry: SL9A1_HUMAN) was used as the query sequence and EcNhaA (PDB entry: 1ZCD [23]) as the template structure with default parameters to generate an additional pairwise alignment.

5.6. Homology Modeling

Modeling of the structure of NHE1 (SwissProt entry: SL9A1_HUMAN), residues 126-505, was based on the template structure of EcNhaA (PDB entry 1ZCD; [23]), using the homology modeling program NEST [22] with default parameters. The final model was based on the pairwise alignments constructed as described in the Results.

5.7. Experimental Procedure

The procedures carried out by Katia Herz (under the supervision of Etana Padan) are described in Appendix III.

5.8. Figures

Figs. 1B, 3B, 4, 5 and 6B were drawn with PyMol (DeLano, W.L. The PyMOL Molecular Graphics System (2002) DeLano Scientific, San Carlos, CA, USA. http://www.pymol.org).

6. Tables

6.1. Table 1. Published Mutations in Human NHE1

Mutations in Human NHE1	Structural Location	Conservation (1−9: Variable-to-Conserved) NHE/EcNhaA	Equivalent Residue in EcNhaA	Notes	Citation
Putatively Essential Residues					
Gly148 (examined in the rat NHE1 – residue G152)	Loop1-2	1/3	Ser31	Moderately important for inhibitor binding. G148A – reduced sensitivity to EIPA by 3-fold.	[164]
Pro153+ Pro154 (examined in the rat NHE1 – Pro157+) Pro158)	Loop1-2	2/1 3/1	Gly36+ Trp36	Moderately important for inhibitor binding. P153S/P154F – reduced sensitivity to EIPA by 7-fold.	[164]
Phe161 (examined also for the hamster NHE1 – F165)	TM2	7/8	Ile63	Important for activity and also for inhibitor binding. Suggested as a pore-lining residue. F161Y – increases the K_i for amiloride by 40-fold and reduces the Na^+ transport rate by 3- to 4-fold. The effect was also shown for F161Y+F164Y (substitutions as in NHE4).	[154, 165]

Phe162 +Ile169 +Ile170	TM2 TM2	7/9 9/9 8/9	Asn64 Phe71+ Phe72	Important for transport and inhibitor binding. F162S – dramatic decrease in affinity for cariporide and amiloride by 1550- and 150-fold, respectively. The mutant also reduces the affinity for sodium. F162S together with I169S or I170T – Normal cation-binding (reversion phenotype), but still interference with the inhibition by cariporide. F162S/I169S/I170T triple mutant – the affinity to sodium is as for wt (reversion phenotype), low affinity to cariporide, and a 4- to 5-fold increase in its Km for lithium. I169S+I170T – Interference with ion-translocation, but no effect on inhibition by cariporide. I169S alone – No effect. I170T alone – No effect.	[166]
Leu163 (examined also for the hamster NHE1 – L167)	TM2	5/9	Asp65	Important for inhibitor binding. L163F – reduces the effectiveness of EIPA and MPA to block transport by ~30-fold, and that of amiloride by 3-fold. Note: the corresponding residue to NHE1's Leu163 in rabbit NHE2 (Leu143) is also involved in amiloride binding [167].	[164, 165]
Pro167	TM2	8/8	Ala69	Important for activity. P167G/C - abolishes NHE activity. P167A - markedly decreases activity. P167G/C/A- expressed at lower levels compared to	[168]

				wt, and a significant portion of P167G and P167C is retained intracellularly.	
Pro168	TM2	7/5	Val70	Important for activity. P168A/C - abolishes activity. P168G - markedly decreases activity. P168 G/C/A -expressed at levels similar to wt and targeted to the plasma membrane.	[168]
Gly174	TM2	8/9	Gly76	Moderately important for inhibitor binding. G174S – shows 3.3-fold decrease in amiloride affinity. G174D – shows 4-fold decrease in amiloride affinity. G174A – no change is amiloride affinity.	[147]
Leu163+ Gly174	TM2	5/9 8/9	Asp65+ Gly76	Important for inhibitor binding. L163F/G174S double mutant – as active as the wt, but reduced affinity for sodium and for amiloride and its derivatives compared to that of the wt. Each of them alone does not affect Na^+ affinity.	[165]
Arg180	TM2	5/9	Glu82	Important for allosteric regulation. R180K − reduces sensitivity to intracellular protons.	[169]
Ile251	Loop 4-5	8/8	Val148	Moderately important for inhibitor binding. I251A − more resistant to EMD87580 by 3-fold compared to wt.	[146]
Leu255	TM5	7/8	Leu152	Important for activity and moderately important for inhibitor binding. L255A - more sensitive to EMD87580 by 2.5-fold.	[146]

				Also show direct effects on activity.	
Ile257	TM5	7/6	Ile154	Important for activity. I257A − shows direct effects on activity.	[146]
Val259	TM5	8/9	Leu156	Important for activity. V259A − shows direct effects on activity.	[146]
Phe260	TM5	8/6	Met157	Important for activity. F260A − shows direct effects on activity.	[146]
Gly261	TM5	9/8	Ala158	Important for activity. G261A − shows direct effects on activity.	[146]
Glu262	TM5	9/8	Leu159	Important for activity. E262Q − abolishes activity. E262D − reduces activity (less than 25% of that of the wt) and also decreases affinity for Li^+. The substitution of aspartate for glutamate, with its smaller side, chain may reduce the ability of the protein to coordinate the smaller lithium ion, while still allowing for coordination of the larger Na^+ ion. E262D/Q − expressed and targeted to the plasma membrane. E262I − abolishes activity; the nature of the defect was not investigated. E262A − low expression and major effects on targeting (found principally in intracellular compartments). However, even after correcting for targeting and expression, the corrected activity of Glu262 was only 52% of control levels, indicating there was still a significant defect in the Na^+/H^+ exchanger	[144, 146, 171]

				activity. Note: the residue that corresponds to E262 in the sod2 transporter from *Schizosaccharomyces pombe* (fission yeast) (Glu173) is also critical for activity [170].	
Leu265	TM5	8/7	Ile162	Moderately important for inhibitor binding. L265A − more sensitive to EMD87580 by 10-fold.	[146]
Asn266	TM5	9/9	Asp163	Important for activity. N266A − shows direct effects on activity.	[146]
Asp267	TM5	9/9	Asp164	Important for activity. D267N − no activity. D267E – active. Note: the residue that corresponds to E262 in the sod2 transporter from *Schizosaccharomyces pombe* (fission yeast) (Asp178) is also critical for activity [172].	[144, 146]
Thr270	TM5	8/9	Ala167	Important for activity. T270A − shows direct effects on activity.	[146]
Val272	TM5	9/4	Ile169	Moderately important for inhibitor binding. V272A - more resistant to EMD87580 by 5-fold. The mutant also shows reduced expression with no effect on activity.	[146]
Leu273	TM5	7/8	Ile170	Moderately important for inhibitor binding. L273A – shows reduced expression, but after correction for protein expression, its activity was greater than that of controls. Also shows increased resistance to EMD87580 by 2.5 fold.	[146]
Arg327	Loop 6-7	4/-	Gap(204-	Important for pH	[169]

			205)	regulation. R327E − significantly shifts the pH set point of NHE1 to the acidic side. R327K – no effect.	
Arg330	Loop6-7	6/-	Gap (204-205)	Important for allosteric regulation. E330M − reduces sensitivity to intracellular protons. R330Q – no effect.	[169]
Glu346 (examined also in the rat NHE1 – E350)	Loop 7-8	7/7	Leu220	Important for activity and for inhibitor binding. E346Q – reduces sensitivity to amiloride and EIPA by 20- and 127-fold, respectively, and also reduces the transport rate. E346/D/N − reduces sensitivity to EIPA by ~500- and ~60-fold, respectively, and also reduces the transport rate. E346D − also reduces the affinity for sodium. This residue has no effect on pH regulation.	[164, 173]
His349	TM8	5/9	Gly223	Moderately important for inhibitor binding. H349Y/F – increases sensitivity to amiloride by ~2-fold. H349G/L – reduces sensitivity to amiloride by ~2-fold. H349S (mimics NHE3) – practically no effects on amiloride sensitivity. H349G/L/Y/F/S – all as active as the wt, with similar affinities for the cations.	[174]
Gly352 (examined in the rat NHE1 – G356)	TM8	8/8	Ala226	Important for activity and inhibitor binding. G356A − reduces the apparent half-maximal inhibition of NHE1 by amiloride, EIPA and	[164]

				HOE694 by ~5-, 32-, and 342-fold, respectively. G352S − decreases EIPA sensitivity by 72-fold. G352D − decreases EIPA sensitivity by 33-fold. G352S/L/D/K – significantly reduces transport. This residue has no effect on pH regulation. Note: the corresponding residue in NHE3 (A305) is also involved in amiloride binding [164]. Also, the corresponding residue in *C. albicans* Cnh1p (Asp267) [175], *S. cerevisiae* Nha1p (Asp267) [176], *Schizosaccharomyces pombe* (Asp267) [170, 172], and *Zygosaccharomyces rouxii* Sod2-22p (Asp266) [177] are crucial for function.	
Leu163+ Gly352 (examined in rat NHE1 L167+ G356)	TM2 TM8	5/9 8/8	Asp65 Ala226	Involved in inhibitor-binding. L163F/G352A – shows 164-fold lower affinity for EIPA compared to wt.	[164]
Glu391	TM9	9/7	Leu264	Important for activity. E391Q – activity is greatly reduced, but not abolished. E391D – active. Note: the corresponding residue in In the *Saccharomyces cerevisiae* Nhx1 (Glu355) is also important for activity [178].	[144]
Cys421	TM10	6/9	Leu296	Important for activity. C421S − small decrease in activity, located in the plasma membrane.	[179]
Arg440	Loop10-11	4/6	Lys315	Important for pH regulation.	[180]

				R440C/D/E/H/L – shift the pH set point of NHE1 to the acidic side with no effect on ion translocation.	
Gly455	TM11	8/7	Cys335	Important for pH regulation. G455C/Q/T/V − shift the pH set point of NHE1 to the alkaline side with no effect on ion translocation. The shift grows larger as the residue becomes bulkier, suggesting a steric hindrance.	[180]
Gly456	TM11	9/9	Gly336	Important for pH regulation. G456C − shifts the pH set point of NHE1 to the alkaline side with no eaffect on ion translocation.	[180]
		Putatively Nonessential Residues			
Cys8	N-terminal	-	-	Essential for localization; nonessential for function C8R/S − intracellular localization (Golgi). However, small amounts of antiporter that reach the membrane are active.	[179]
Asn75	N-terminal	-	-	Glycosylation site. N75D – active as the wt.	[181]
Cys113	N-terminal	-	-	C113S − as active as the WT, located in the plasma membrane.	[179]
His120	N-terminal	-	-	H120G – as active as the wt.	[171, 174]
Cys133	TM1	7/5	Ile16	C133S − as active as the WT, located in the plasma membrane.	[179]
Pro178	TM2	7/9	Lys80	P178A − as active as the wt, located in the plasma membrane, expressed at levels similar to wt.	[168]

Glu184	Loop 2-3	4/9	Gly86	E184H − Note: Mutation in the corresponding position in sod2 (His98) also failed to change activity [172, 182]	[169]
Cys212	Loop 3-4	2	Gap(116-117)	C212S − as active as the wt. Located in the plasma membrane.	[179]
Asp238	TM4	9/9	D133	D238N − as active as the wt. Note: the corresponding residue in the sod2 transporter from *Schizosaccharomyces pombe* (fission yeast) (Asp145) is critical for activity [170, 172].	[144]
Pro239	TM4	9/9	I134	P239A − as active as the wt. Note: the corresponding residue in the Sod2-22 transporter from *Zygosaccharomyces rouxii* (Pro145) is critical for activity [183].	[144]
Glu248	TM4	6/8	Leu143	E248Q – as active as wt.	[169]
His250	Loop4-5	5/7	Arg147	Preliminary data suggest that it is not involved in exchanger function.	[171, 174]
Leu254	TM5	6/7	Ala151	L254A – shows reduced expression but no effect on activity.	[146]
His256	TM5	6/9	Lys153	H256A – shows reduced expression but no effect on activity.	[146]
Ser263	TM5	9/9	Ala160	S263A − as active as the wt.	[144]
Val269	TM5	8/7	Gly166	V269A – active.	[146]
Val271	TM5	8/9	Ile168	V271A – shows reduced expression but no effect on activity.	[146]
His325	Loop6-7	4/-	Gap(204-205)	H325C − as active as the wt.	[171, 184]

Ser359	TM8	7/6	Val233	S359A − as active as the wt.	[144]
Asn370	Loop 8-9	6/1	His243	N370D – as active as the wt.	[181]
Ser387+ Ser388	TM9	7/6 6/6	Ala260+ Tyr261	S387A/S388A – as active as the wt.	[144]
Ser390	TM9	8/7	Ile263	S390A – as active as the wt.	[144]
Thr392	TM9	7/9	Pro265	T392V – as active as the wt.	[144]
Ser401	Loop9-10	7/5	Ser275	S401A – as active as the wt.	[144]
Thr402	Loop9-10	6/4	Leu276	T402V – as active as the wt.	[144]
Ser406	Loop9-10	1/2	Thr280	S406A – as active as the wt.	[144]
Asn410	Loop9-10	1/1	Thr285	N410D– as active as the wt.	[181]
Lys438	Loop 10-11	1/6	Arg313	K438E– as active as the wt.	[169]
Lys443	Loop 10-11	1/1	His318	K443E– as active as the wt.	[169]
Cys477	Loop 11-12	6/2	D354	Essential for localization; nonessential for function. C477S − intracellular localization (Golgi). However, small amounts of antiporter that reach the membrane are active.	[179]

Table 1: The mutations are divided into putatively essential and nonessential sites. The locations of the mutations on the predicted topology suggested in this paper and their corresponding numbers in EcNhaA are indicated. Conservation scores for NHE1 and EcNhaA are according to the ConSurf 1−9 (variable-to-conserved) scale, calculated using 305 NHE1-related Na^+/H^+ exchanger homologous sequences or 94 bacterial NhaA Na^+/H^+ antiporters, respectively, as described in Methods in chapter 2. EMD87580 is an amiloride analog. MPA: N^5-methyl-N^5-propylamiloride; EIPA: 5-(*N*-ethyl-*N*-isopropyl) amiloride.

6.2. Table 2. Published Mutations in Eukaryotic Na^+/H^+ Exchangers

Mutations	Structural Location	Conservation (1−9: Variable-to-Conserved) NHE/EcNhaA	Corresponding Residue in NHE1/EcNhaA	Notes	Citation
Putatively Essential Residues					
Leu143 In rabbit NHE2 (SL9A2_RABIT)	TM2	5/9	Leu163/Asp65	Important for inhibitor binding. Note: the corresponding position in NHE1 is also involved in amiloride binding [165].	[167]
Thr141 In *Zygosaccharomyces rouxii* Sod2-22p (Q9UUT4_ZYGRO)	TM4	8/7	Ser235 /Ala130	Important for activity. T141S produces a broadened cation selectivity of the antiporter for K^+, in addition to Na^+ and Li^+.	[185]
Asp145 In *Schizosaccharomyces pombe* (fission yeast) sod2 (NAH_SCHPO)	TM 4	9/9	Asp238/ Asp133	Important for activity. Note: in NHE1, mutation of D238N is not crucial for function [144].	[170, 172]
P145 In *Zygosaccharomyces rouxii* Sod2-22p (Q9UUT4_ZYGRO)	TM4	9/9	Pro239/Ile134	Important for activity. P145S/T − decreases the antiporter transport activity for both Na^+ and Li^+, yet enables ZrSod2-22p to transport K^+. P145D/K − abolishes activity. P145G − extremely low activity for exchanging Na^+. Note: the corresponding residues in *S. cerevisiae* Nha1 antiporter (Pro146) are also critical for	[183]

				function [183]. However, mutation in the equivalent position in NHE1(P239A) is not crucial for activity [144].	
Ser150 In *Zygosaccharomyces rouxii* Sod2-22p (Q9UUT4_ZYGRO)	TM4	8/8	Ala244/ Gly139	Important for activity. S150T produces a broadened cation selectivity of the antiporter for K^+, in addition to Na^+ and Li^+. S150K/R – abolishes activity. S150V/D − no significant change in substrate specificity.	[185]
Glu173 In *Schizosaccharomyces pombe* (fission yeast) sod2 (NAH_SCHPO)	TM 5	9/8	Glu262/Leu159	Important for activity. Note: the corresponding residue in NHE1 is also crucial for function [144].	[170]
Asp178 In *Schizosaccharomyces pombe* (fission yeast) sod2 (NAH_SCHPO)	TM 5	9/9	Asp267/ Asp164	Important for activity. Note: the corresponding residue in NHE1 is also crucial for function [144].	[172]
Asp241 In *Saccharomyces cerevisiae* Nha1p (NAH1_YEAST)	loop 6-7	6/-	Ile326/gap	Intermediate importance for activity. D241N − little effect on Na^+ efflux but significantly reduces K^+ efflux. Note: the corresponding residue in sod2 *Schizosaccharomyces pombe* (Asp241)	[176]

				displays intermediate to insignificant importance for the function of the transporter [170, 172, 182].	
Asp265 In *Zygosaccharomyces rouxii* Sod2-22p (Q9UUT4_ZYGRO)	TM8	9/9	Ser351/His225	Important for activity. D265V – not active. Note: the corresponding residues in *C. albicans* Cnh1p (Asp266) [175], *S. cerevisiae* Nha1p (Asp266) [176] and *Schizosaccharomyces pombe* (Asp266) [170, 172] are also crucial for function.	[177]
Asp266 In *Zygosaccharomyces rouxii* Sod2-22p (Q9UUT4_ZYGRO)	TM8	8/8	Gly352/ Ala226	Important for activity. D266V – not active. Note: the corresponding residues in rat NHE1 (G356) and also in NHE3 (A305) are involved in amiloride binding [164]. Also, the corresponding residues in *C. albicans* Cnh1p (Asp267) [175], *S. cerevisiae* Nha1p (Asp267) [176] and *Schizosaccharomyces pombe* (Asp267) [170, 172] are crucial for function.	[177]
Glu355 In the *Saccharomyces cerevisiae* Nhx1 (NAH2_YEAST)	TM9	9/7	Glu391/Leu264	Important for activity. E355A – sever impairment of growth. E355Q- partial impairment of growth. Note: the corresponding residue in NHE1 is also important to activity [144].	[178]
His367 In	TM 11	7/7	Tyr454/Leu334	Important for activity and pH regulation.	[172, 182]

Schizosaccharomyces pombe (fission yeast) sod2 (NAH_SCHPO)				H367/A - not active. H367D - pH shift to more alkaline pH. Note: the corresponding residue in NHE1 is important for cellular localization [184].	
		Putatively Nonessential Residues			
His67 In *Schizosaccharomyces pombe* (fission yeast) sod2 (NAH_SCHPO)	TM2	5/6	Gln157/Met59	Mutation does not significantly impair proton translocation.	[182]
His98 In *Schizosaccharomyces pombe* (fission yeast) sod2 (NAH_SCHPO)	loop 2-3	4/9	Glu184/Gly86	Mutation does not significantly impair proton translocation. Note: mutation in the corresponding position in NHE1 also fails to change activity [169].	[172, 182]
His233 In *Schizosaccharomyces pombe* (fission yeast) sod2 (NAH_SCHPO)	loop 6-7	3/-	Phe322/gap	Mutation does not significantly impair proton translocation.	[172, 182]
His424 In *Schizosaccharomyces pombe* (fission yeast) sod2 (NAH_SCHPO)	TM12	9/8	Gln495/ Ser372	Mutation does not significantly impair proton translocation.	[182]
His429 In *Schizosaccharomyces pombe* (fission	TM12	6/1	Arg500/ Tyr377	Mutation does not significantly impair proton translocation.	[182]

yeast) sod2 (NAH_SCHPO)					

Table 2: The mutations are divided into putatively essential and nonessential sites. The locations of the mutations on the corresponding positions in the predicted topology suggested in this paper for NHE1 and their corresponding numbers in EcNhaA are indicated. Conservation scores are according to the ConSurf 1−9 (variable-to-conserved) scale, calculated using 305 NHE1-related Na^+/H^+ exchanger homologous sequences or 94 bacterial NhaA Na^+/H^+ antiporters, as described in Methods in chapter 2.

6.3. Table 3. AP sensitivity of EcNhaA mutants

	Location of the Mutation	**NaCl IC_{50} (µM AP)**	**LiCl IC_{50} (µM AP)**
Wild-type		1.7	2.2
W62C	TM2	1.2	1.5
F71C	TM2	**0.3**	0.8*
F72C	TM2	1.8	2.6
G76C	TM2	1.6	1.2
N64C	TM2	**0.5**	**17**
H225R	TM8	1*	2.3
H225Q	TM8	0.8	**7.8**

Table 3: For calculation of the IC_{50} values of AP inhibition, the activity of the antiporter in percent dequenching (100% corresponds to the activity in the absence of AP) was plotted versus different AP concentration as previously described [134]. The Na^+/H^+ or Li^+/H^+ antiporter activity was measured at various ion concentrations around the apparent K_M at the pH of maximal activity (pH 8.5 or pH 7.5 *). IC_{50} values of AP inhibition in mutants that are different by more than 3-fold than wild-type are bolded.

CONCLUDING REMARKS

During my doctoral work, I came to the understanding that using a data-integration approach has the potential to go a long way in providing new frameworks for mechanistic understanding of central processes in biology. Most prominently, I greatly appreciate the significance of conducting computational studies that are coupled to experiments. My work has been focused on the examination of two transmembrane protein families, both essential to cellular functions and also play a role in terminal diseases. Therefore, the understanding of their regulatory mechanisms at the molecular level is imperative and directly related to developing therapies.

My research on the ErbB family employed structural and phylogenetic analyses, integrated with published experimental data, to explore the mechanism of ErbB regulation at the molecular level. This research presents a comprehensive view of the available data on these proteins, which has now reached a crucial stage, where the experimental data and models are on the point of converging in support of a unified mechanism of receptor activation. My research also discusses the significance of this mechanism for pathological conditions. Finally, I studied the molecular interpretation of the effect of cancer-causing mutations and their implication on therapy with the anti-ErbB drugs.

My work on the Na^+/H^+ exchangers resulted in a model structure for NHE1. This model provided a framework for integrating phylogenetic and biochemical data to suggest a scheme for the ion-translocation mechanism at the molecular level. It is likely that my model structure will lead to revision of the field, because it provides the means for formulating hypotheses and testing them empirically, e.g., by using mutations as done by my collaborators. It might also pave the way to structure-based drug design of additional NHE1 inhibitors of clinical benefit.

REFERENCES

1. Petrey, D. and B. Honig, Protein structure prediction: inroads to biology. *Mol. Cell* (2005) **20**: 811-819.
2. Huse, M. and J. Kuriyan, The conformational plasticity of protein kinases. *Cell* (2002) **109**: 275-82.
3. Glaser, F., T. Pupko, I. Paz, R.E. Bell, D. Bechor-Shental, E. Martz and N. Ben-Tal, ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. *Bioinformatics* (2003) **19**: 163-4.
4. Pupko, T., R.E. Bell, I. Mayrose, F. Glaser and N. Ben-Tal, Rate4Site: an algorithmic tool for the identification of functional regions in proteins by surface mapping of evolutionary determinants within their homologues. *Bioinformatics* (2002) **18**: S71-7.
5. Landau, M., I. Mayrose, Y. Rosenberg, F. Glaser, E. Martz, T. Pupko and N. Ben-Tal, ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res.* (2005) **33**: W299-302.
6. Nimrod, G., F. Glaser, D. Steinberg, N. Ben-Tal and T. Pupko, In silico identification of functional regions in proteins. *Bioinformatics* (2005) **21**: i328-37.
7. Laskowski, R.A., J.D. Watson and J.M. Thornton, ProFunc: a server for predicting protein function from 3D structure. *Nucleic Acids Res.* (2005) **33**: W89-93.
8. Ferrer-Costa, C., M. Orozco and X. de la Cruz, Characterization of disease-associated single amino acid polymorphisms in terms of sequence and structure properties. *J. Mol. Biol.* (2002) **315**: 771-86.
9. Sridharan, S., A. Nicholls and B. Honig, A new vertex algorithm to calculate solvent accessible surface area. *Biophys. J.* (1992) **61**: A174.
10. Berezin, C., F. Glaser, J. Rosenberg, I. Paz, T. Pupko, P. Fariselli, R. Casadio and N. Ben-Tal, ConSeq: the identification of functionally and structurally important residues in protein sequences. *Bioinformatics* (2004) **20**: 1322-4.
11. Ashur-Fabian, O., A. Avivi, L. Trakhtenbrot, K. Adamsky, M. Cohen, G. Kajakaro, A. Joel, N. Amariglio, E. Nevo and G. Rechavi, Evolution of p53 in hypoxia-stressed Spalax mimics human tumor mutation. *Proc. Natl. Acad. Sci. USA* (2004) **101**: 12236-41.

12. Fromovich-Amit, Y., A. Zivelin, N. Rosenberg, H. Tamary, M. Landau and U. Seligsohn, Characterization of mutations causing factor VII deficiency in 61 unrelated Israeli patients. *J. Thromb. Haemost.* (2004) **2**: 1774-81.

13. Rosenberg, N., M. Landau, J. Luboshitz, G. Rechavi and U. Seligsohn, A novel PHE171CYS mutation in integrin αIIB causes Glanzmann Thrombasthenia by abrogating αIIbβ3 complex formation. *J. Thromb. Haemost.* (2004) **2**: 1167-75.

14. Vysokovsky, A., R. Saxena, M. Landau, A. Zivelin, R. Eskaraev, N. Rosenberg, U. Seligsohn and A. Inbal, Seven novel mutations in the factor XIII A-subunit gene causing hereditary factor XIII deficiency in 10 unrelated families. *J. Thromb. Haemost.* (2004) **2**: 1790-7.

15. Zivelin, A., T. Ogawa, S. Bulvik, M. Landau, J.R. Toomey, J. Lane, U. Seligsohn and D. Gailani, Severe factor XI deficiency caused by a Gly555 to Glu mutation (factor XI-Glu555): a cross-reactive material positive variant defective in factor IX activation. *J. Thromb. Haemost.* (2004) **2**: 1782-9.

16. Fromovich-Amit, Y., A. Zivelin, N. Rosenberg, M. Landau, J.P. Rosa and U. Seligsohn, Of four mutations in the factor VII gene in Tunisian patients, one novel mutation (Ser339Phe) in three unrelated families abrogates factor X activation. *Blood Coagul. Fibrinolysis* (2005) **16**: 369-374.

17. Rosenberg, N., H. Hauschner, H. Peretz, R. Mor-Cohen, M. Landau, B. Shenkman, G. Kenet, B.S. Coller, A.A. Awidi and U. Seligsohn, A 13-bp deletion in alphaIIb gene is a founder mutation that predominates in Palestinian-Arab patients with Glanzmann thrombasthenia. *J. Thromb. Haemost.* (2005) **3**: 2764-72.

18. Peretz, H., N. Rosenberg, M. Landau, S. Usher, E.J. Nelson, R. Mor-Cohen, D.L. French, B.W. Mitchell, S.C. Nair, M. Chandy, *et al.*, Molecular diversity of Glanzmann thrombasthenia in southern India: new insights into mRNA splicing and structure-function correlations of alphaIIbbeta3 integrin (ITGA2B, ITGB3). *Hum. Mutat.* (2006) **27**: 359-69.

19. Rosenberg, N., S. Lalezari, M. Landau, B. Shenkman, U. Seligsohn and S. Izraeli, Trp207Gly in platelet glycoprotein Ibalpha is a novel mutation that disrupts the connection between the leucine-rich repeat domain and the disulfide loop structure and causes Bernard-Soulier syndrome. *J. Thromb. Haemost.* (2007) **5**: 378-86.

20. Landau, M., S.J. Fleishman and N. Ben-Tal, A putative mechanism for downregulation of the catalytic activity of the EGF receptor via direct contact between its kinase and C-terminal domains. *Structure* (2004) **12**: 2265-75.
21. Fleishman, S.J. and N. Ben-Tal, Progress in structure prediction of alpha-helical membrane proteins. *Curr. Opin. Struc. Biol.* (2006) **16**: 496-504.
22. Petrey, D., Z. Xiang, C.L. Tang, L. Xie, M. Gimpelev, T. Mitros, C.S. Soto, S. Goldsmith-Fischman, A. Kernytsky, A. Schlessinger, *et al.*, Using multiple structure alignments, fast model building, and energetic analysis in fold recognition and homology modeling. *Proteins* (2003) **53**: 430-435.
23. Hunte, C., E. Screpanti, M. Venturi, A. Rimon, E. Padan and H. Michel, Structure of a Na+/H+ antiporter and insights into mechanism of action and regulation by pH. *Nature* (2005) **435**: 1197-1202.
24. Adamian, L. and J. Liang, Prediction of transmembrane helix orientation in polytopic membrane proteins. *BMC. Struct. Biol.* (2006) **6**: 13.
25. Fleishman, S.J., V.M. Unger, M. Yeager and N. Ben-Tal, A C[alpha] model for the transmembrane [alpha] helices of gap junction intercellular channels. *Mol. Cell* (2004) **15**: 879-888.
26. Baldwin, J.M., G.F.X. Schertler and V.M. Unger, An alpha-carbon template for the transmembrane helices in the rhodopsin family of G-protein-coupled receptors. *J. Mol. Biol.* (1997) **272**: 144-164.
27. Briggs, J.A.G., J. Torres and I. Arkin, A new method to model membrane protein structure based on silent amino acid substitutions. *Proteins* (2001) **44**: 370-375.
28. Hurwitz, N., M. Pellegrini-Calace and D.T. Jones, Towards genome-scale structure prediction for transmembrane proteins. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* (2006) **361**: 465-75.
29. Fleishman, S.J., V.M. Unger and N. Ben-Tal, Transmembrane protein structures without X-rays. *Trends Biochem. Sci.* (2006) **31**: 106-13.
30. Stein, R. and J. Staros, Insights into the evolution of the ErbB receptor family and their ligands from sequence analysis. *BMC Evol. Biol.* (2006) **6**: 79.
31. Yarden, Y. and M.X. Sliwkowski, Untangling the ErbB signalling network. *Nat. Rev. Mol. Cell. Biol.* (2001) **2**: 127-37.
32. Citri, A. and Y. Yarden, EGF-ERBB signalling: towards the systems level. *Nat. Rev. Mol. Cell. Biol.* (2006) **7**: 505-16.

33. Linggi, B. and G. Carpenter, ErbB receptors: new insights on mechanisms and biology. *Trends Cell Biol.* (2006) **16**: 649-656.
34. Schlessinger, J., Cell signaling by receptor tyrosine kinases. *Cell* (2000) **103**: 211-25.
35. Ozcan, F., P. Klein, M.A. Lemmon, I. Lax and J. Schlessinger, On the nature of low- and high-affinity EGF receptors on living cells. *Proc. Natl. Acad. Sci. U.S.A.* (2006) **103**: 5735-40.
36. Warren, C.M. and R. Landgraf, Signaling through ERBB receptors: Multiple layers of diversity and control. *Cell. Signal.* (2006) **18**: 923-933.
37. Zhang, X., J. Gureasko, K. Shen, P.A. Cole and J. Kuriyan, An allosteric mechanism for activation of the kinase domain of epidermal growth factor receptor. *Cell* (2006) **125**: 1137-49.
38. Stamos, J., M.X. Sliwkowski and C. Eigenbrot, Structure of the epidermal growth factor receptor kinase domain alone and in complex with a 4-anilinoquinazoline inhibitor. *J. Biol. Chem.* (2002) **277**: 46265-72.
39. Fleishman, S.J., J. Schlessinger and N. Ben-Tal, A putative molecular-activation switch in the transmembrane domain of erbB2. *Proc. Natl. Acad. Sci. U.S.A.* (2002) **99**: 15937-40.
40. Li, S., K.R. Schmitz, P.D. Jeffrey, J.J. Wiltzius, P. Kussie and K.M. Ferguson, Structural basis for inhibition of the epidermal growth factor receptor by cetuximab. *Cancer cell* (2005) **7**: 301-11.
41. Ogiso, H., R. Ishitani, O. Nureki, S. Fukai, M. Yamanaka, J.H. Kim, K. Saito, A. Sakamoto, M. Inoue, M. Shirouzu, *et al.*, Crystal structure of the complex of human epidermal growth factor and receptor extracellular domains. *Cell* (2002) **110**: 775-87.
42. Kani, K., C.M. Warren, C.S. Kaddis, J.A. Loo and R. Landgraf, Oligomers of ERBB3 havetwo distinct interfaces that differ in their sensitivity to disruption by heregulin. *J. Biol. Chem.* (2005) **280**: 8238-8247.
43. Nagy, P., G. Vereb, Z. Sebestyen, G. Horvath, S.J. Lockett, S. Damjanovich, J.W. Park, T.M. Jovin and J. Szollosi, Lipid rafts and the local density of ErbB proteins influence the biological role of homo- and heteroassociations of ErbB2. *J. Cell Sci.* (2002) **115**: 4251-62.
44. Yu, X., K.D. Sharma, T. Takahashi, R. Iwamoto and E. Mekada, Ligand-independent dimer formation of epidermal growth factor receptor (EGFR) is a step separable from ligand-induced EGFR signaling. *Mol. Biol. Cell* (2002) **13**: 2547-57.

45. Chantry, A., The kinase domain and membrane localization determine intracellular interactions between epidermal growth factor receptors. *J. Biol. Chem.* (1995) **270**: 3068-73.

46. Martin-Fernandez, M., D.T. Clarke, M.J. Tobin, S.V. Jones and G.R. Jones, Preformed oligomeric epidermal growth factor receptors undergo an ectodomain structure change during signaling. *Biophys. J.* (2002) **82**: 2415-2427.

47. Mendrola, J.M., M.B. Berger, M.C. King and M.A. Lemmon, The single transmembrane domains of ErbB receptors self-associate in cell membranes. *J. Biol. Chem.* (2002) **277**: 4704-4712.

48. Gadella, T.W., Jr. and T.M. Jovin, Oligomerization of epidermal growth factor receptors on A431 cells studied by time-resolved fluorescence imaging microscopy. A stereochemical model for tyrosine kinase receptor activation. *J. Cell Biol.* (1995) **129**: 1543-58.

49. Moriki, T., H. Maruyama and I.N. Maruyama, Activation of preformed EGF receptor dimers by ligand-induced rotation of the transmembrane domain. *J. Mol. Biol.* (2001) **311**: 1011-26.

50. Cadena, D.L., C.L. Chan and G.N. Gill, The intracellular tyrosine kinase domain of the epidermal growth factor receptor undergoes a conformational change upon autophosphorylation. *J. Biol. Chem.* (1994) **269**: 260-5.

51. Teramura, Y., J. Ichinose, H. Takagi, K. Nishida, T. Yanagida and Y. Sako, Single-molecule analysis of epidermal growth factor binding on the surface of living cells. *EMBO J.* (2006) **25**: 4215-22.

52. Kornev, A.P., N.M. Haste, S.S. Taylor and L.F. Ten Eyck, Surface comparison of active and inactive protein kinases identifies a conserved activation mechanism. *Proc. Natl. Acad. Sci. U.S.A.* (2006) **103**: 17783-17788.

53. Pellicena, P. and J. Kuriyan, Protein-protein interactions in the allosteric regulation of protein kinases. *Curr. Opin. Struc. Biol.* (2006) **16**: 702-709.

54. Gotoh, N., A. Tojo, M. Hino, Y. Yazaki and M. Shibuya, A highly conserved tyrosine residue at codon 845 within the kinase domain is not required for the transforming activity of human epidermal growth factor receptor. *Biochem. Biophys. Res. Commun.* (1992) **186**: 768-74.

55. Wood, E.R., A.T. Truesdale, O.B. McDonald, D. Yuan, A. Hassell, S.H. Dickerson, B. Ellis, C. Pennisi, E. Horne, K. Lackey, *et al.*, A unique structure for epidermal growth factor receptor bound to GW572016 (Lapatinib): relationships among protein

conformation, inhibitor off-rate, and receptor activity in tumor cells. *Cancer Res.* (2004) **64**: 6652-6659.

56. McLaughlin, S., S.O. Smith, M.J. Hayman and D. Murray, An electrostatic engine model for autoinhibition and activation of the epidermal growth factor receptor (EGFR/ErbB) family. *J. Gen. Physiol.* (2005) **126**: 41-53.

57. Choi, S.H., J.M. Mendrola and M.A. Lemmon, EGF-independent activation of cell-surface EGF receptors harboring mutations found in gefitinib-sensitive lung cancer. *Oncogene* (2007) **26**: 1567-76.

58. Chang, C.M., H.K. Shu, L. Ravi, R.J. Pelley, H. Shu and H.J. Kung, A minor tyrosine phosphorylation site located within the CAIN domain plays a critical role in regulating tissue-specific transformation by erbB kinase. *J. Virol.* (1995) **69**: 1172-80.

59. Massoglia, S., A. Gray, T.J. Dull, S. Munemitsu, H.J. Kun, J. Schlessinger and A. Ullrich, Epidermal growth factor receptor cytoplasmic domain mutations trigger ligand-independent transformation. *Mol. Cell Biol.* (1990) **10**: 3048-55.

60. Riedel, H., J. Schlessinger and A. Ullrich, A chimeric, ligand-binding v-erbB/EGF receptor retains transforming potential. *Science* (1987) **236**: 197-200.

61. Boerner, J.L., A. Danielsen and N.J. Maihle, Ligand-independent oncogenic signaling by the epidermal growth factor receptor: v-ErbB as a paradigm. *Exp. Cell Res.* (2003) **284**: 111-121.

62. Gamett, D.C., S.E. Tracy and H.L. Robinson, Differences in sequences encoding the carboxyl-terminal domain of the epidermal growth factor receptor correlate with differences in the disease potential of viral erbB genes. *Proc. Natl. Acad. Sci. U.S.A.* (1986) **83**: 6053-7.

63. Pelley, R.J., N.J. Maihle, C. Boerkoel, H.K. Shu, T.H. Carter, C. Moscovici and H.J. Kung, Disease tropism of c-erbB: effects of carboxyl-terminal tyrosine and internal mutations on tissue-specific transformation. *Proc. Natl. Acad. Sci. U.S.A.* (1989) **86**: 7164-8.

64. Raines, M.A., N.J. Maihle, C. Moscovici, M.G. Moscovici and H.J. Kung, Molecular characterization of three erbB transducing viruses generated during avian leukosis virus-induced erythroleukemia: extensive internal deletion near the kinase domain activates the fibrosarcoma- and hemangioma-inducing potentials of erbB. *J. Virol.* (1988) **62**: 2444-52.

65. Robinson, H.L., S.E. Tracy, N. Nair, C. Taglienti-Sian and D.C. Gamett, Characterization of an angiosarcoma-inducing mutation in the erbB oncogene. *Oncogene* (1992) **7**: 2025-30.

66. Jeffrey, P.D., A.A. Russo, K. Polyak, E. Gibbs, J. Hurwitz, J. Massague and N.P. Pavletich, Mechanism of CDK activation revealed by the structure of a cyclinA-CDK2 complex. *Nature* (1995) **376**: 313-20.

67. Aifa, S., J. Aydin, G. Nordvall, I. Lundstrom, S.P.S. Svensson and O. Hermanson, A basic peptide within the juxtamembrane region is required for EGF receptor dimerization. *Exp. Cell Res.* (2005) **302**: 108-114.

68. Sato, T., P. Pallavi, U. Golebiewska, S. McLaughlin and S.O. Smith, Structure of the membrane reconstituted transmembrane-juxtamembrane peptide EGFR(622-660) and its interaction with Ca2+/calmodulin. *Biochemistry* (2006) **45**: 12704-14.

69. Enosh, A., S.J. Fleishman, N. Ben-Tal and D. Halperin, Prediction and simulation of motion in pairs of transmembrane alpha-helices. *Bioinformatics* (2007) **23**: e212-8.

70. Jiang, G. and T. Hunter, Receptor signaling: when dimerization is not enough. *Curr. Biol.* (1999) **9**: R568-71.

71. Cho, H.S. and D.J. Leahy, Structure of the extracellular region of HER3 reveals an interdomain tether. *Science* (2002) **297**: 1330-3.

72. Bouyain, S., P.A. Longo, S. Li, K.M. Ferguson and D.J. Leahy, The extracellular region of ErbB4 adopts a tethered conformation in the absence of ligand. *Proc. Natl. Acad. Sci. U.S.A.* (2005) **102**: 15024-9.

73. Garrett, T.P., N.M. McKern, M. Lou, T.C. Elleman, T.E. Adams, G.O. Lovrecz, H.J. Zhu, F. Walker, M.J. Frenkel, P.A. Hoyne, *et al.*, Crystal structure of a truncated epidermal growth factor receptor extracellular domain bound to transforming growth factor alpha. *Cell* (2002) **110**: 763-73.

74. Garrett, T.P., N.M. McKern, M. Lou, T.C. Elleman, T.E. Adams, G.O. Lovrecz, M. Kofler, R.N. Jorissen, E.C. Nice, A.W. Burgess, *et al.*, The crystal structure of a truncated ErbB2 ectodomain reveals an active conformation, poised to interact with other ErbB receptors. *Mol. Cell* (2003) **11**: 495-505.

75. Cho, H.S., K. Mason, K.X. Ramyar, A.M. Stanley, S.B. Gabelli, D.W. Denney, Jr. and D.J. Leahy, Structure of the extracellular region of HER2 alone and in complex with the Herceptin Fab. *Nature* (2003) **421**: 756-60.

76. Burgess, A.W., H.S. Cho, C. Eigenbrot, K.M. Ferguson, T.P. Garrett, D.J. Leahy, M.A. Lemmon, M.X. Sliwkowski, C.W. Ward and S. Yokoyama, An open-and-shut

case? Recent insights into the activation of EGF/ErbB receptors. *Mol. Cell* (2003) **12**: 541-52.

77. Leahy, D.J., Structure and function of the epidermal growth factor (EGF/ErbB) family of receptors. *Adv. Protein Chem.* (2004) **68**: 1-27.

78. Ferguson, K.M., M.B. Berger, J.M. Mendrola, H.S. Cho, D.J. Leahy and M.A. Lemmon, EGF activates its receptor by removing interactions that autoinhibit ectodomain dimerization. *Mol. Cell* (2003) **11**: 507-17.

79. Mattoon, D., P. Klein, M.A. Lemmon, I. Lax and J. Schlessinger, The tethered configuration of the EGF receptor extracellular domain exerts only a limited control of receptor function. *Proc. Natl. Acad. Sci. U.S.A.* (2004) **101**: 923-8.

80. Klein, P., D. Mattoon, M.A. Lemmon and J. Schlessinger, A structure-based model for ligand binding and dimerization of EGF receptors. *Proc. Natl. Acad. Sci. U.S.A.* (2004) **101**: 929-34.

81. Walker, F. and A.W. Burgess, Reconstitution of the high affinity epidermal growth factor receptor on cell-free membranes after transmodulation by platelet-derived growth factor. *J. Biol. Chem.* (1991) **266**: 2746-2752.

82. Holbrook, M.R., L.L. Slakey and D.J. Gross, Thermodynamic mixing of molecular states of the epidermal growth factor receptor modulates macroscopic ligand binding affinity. *Biochem. J.* (2000) **352**: 99-108.

83. Van der Heyden, M.A.G., M. Nievers, A.J. Verkleij, J. Boonstra and P.M.P. Van Bergen en Henegouwen, Identification of an intracellular domain of the EGF receptor required for high-affinity binding of EGF. *FEBS Lett.* (1997) **410**: 265-268.

84. Li, E. and K. Hristova, Role of receptor tyrosine kinase transmembrane domains in cell signaling and human pathologies. *Biochemistry* (2006) **45**: 6241-6251.

85. Livneh, E., R. Prywes, O. Kashles, N. Reiss, I. Sasson, Y. Mory, A. Ullrich and J. Schlessinger, Reconstitution of human epidermal growth factor receptors and its deletion mutants in cultured hamster cells. *J. Biol. Chem.* (1986) **261**: 12490-12497.

86. Citri, A., K.B. Skaria and Y. Yarden, The deaf and the dumb: the biology of ErbB-2 and ErbB-3. *Exp. Cell Res.* (2003) **284**: 54-65.

87. Berger, M.B., J.M. Mendrola and M.A. Lemmon, ErbB3/HER3 does not homodimerize upon neuregulin binding at the cell surface. *FEBS Lett.* (2004) **569**: 332-6.

88. Gillgrass, A., R.D. Cardiff, N. Sharan, S. Kannan and W.J. Muller, Epidermal growth factor receptor-dependent activation of Gab1 is involved in ErbB-2-mediated mammary tumor progression. *Oncogene* (2003) **22**: 9151-9155.

89. DiGiovanna, M.P., D.F. Stern, S.M. Edgerton, S.G. Whalen, D. Moore, 2nd and A.D. Thor, Relationship of epidermal growth factor receptor expression to ErbB-2 signaling activity and prognosis in breast cancer patients. *J. Clin. Oncol.* (2005) **23**: 1152-1160.

90. Lenferink, A.E.G., J.F. Simpson, L.K. Shawver, R.J. Coffey, J.T. Forbes and C.L. Arteaga, Blockade of the epidermal growth factor receptor tyrosine kinase suppresses tumorigenesis in MMTV/Neu + MMTV/TGF-alpha bigenic mice. *Proc. Natl. Acad. Sci. U.S.A.* (2000) **97**: 9609-9614.

91. Wang, S.E., A. Narasanna, M. Perez-Torres, B. Xiang, F.Y. Wu, S. Yang, G. Carpenter, A.F. Gazdar, S.K. Muthuswamy and C.L. Arteaga, HER2 kinase domain mutation results in constitutive phosphorylation and activation of HER2 and EGFR and resistance to EGFR tyrosine kinase inhibitors. *Cancer cell* (2006) **10**: 25-38.

92. Citri, A., J. Gan, Y. Mosesson, G. Vereb, J. Szollosi and Y. Yarden, Hsp90 restrains ErbB-2/HER2 signalling by limiting heterodimer formation. *EMBO rep.* (2004) **5**: 1165-70.

93. Blume-Jensen, P. and T. Hunter, Oncogenic kinase signalling. *Nature* (2001) **411**: 355-65.

94. Chen, Y.-R., Y.-N. Fu, C.-H. Lin, S.-T. Yang, S.-F. Hu, Y.-T. Chen, S.-F. Tsai and S.-F. Huang, Distinctive activation patterns in constitutively active and gefitinib-sensitive EGFR mutants. *Oncogene* (2005) **25**: 1205-1215.

95. Willmore-Payne, C., J.A. Holden and L.J. Layfield, Detection of epidermal growth factor receptor and human epidermal growth factor receptor 2 activating mutations in lung adenocarcinoma by high-resolution melting amplicon analysis: correlation with gene copy number, protein expression, and hormone receptor expression. *Hum. Pathol.* (2006) **37**: 755-63.

96. Hynes, N.E. and T. Schlange, Targeting ADAMS and ERBBs in lung cancer. *Cancer cell* (2006) **10**: 7-11.

97. Stephens, P., C. Hunter, G. Bignell, S. Edkins, H. Davies, J. Teague, C. Stevens, S. O'Meara, R. Smith, A. Parker, *et al.*, Lung cancer: Intragenic ERBB2 kinase mutations in tumours. *Nature* (2004) **431**: 525-526.

98. Dancey, J.E., Predictive factors for epidermal growth factor receptor inhibitors—The bull's-eye hits the arrow. *Cancer cell* (2004) **5**: 411-415.

99. Greulich, H., T.-H. Chen, W. Feng, P.A. Jänne, J.V. Alvarez, M. Zappaterra, S.E. Bulmer, D.A. Frank, W.C. Hahn, W.R. Sellers, *et al.*, Oncogenic transformation by inhibitor-sensitive and -resistant EGFR mutants. *PLoS Med.* (2005) **2**: e313.

100. Lynch, T.J., D.W. Bell, R. Sordella, S. Gurubhagavatula, R.A. Okimoto, B.W. Brannigan, P.L. Harris, S.M. Haserlat, J.G. Supko, F.G. Haluska, *et al.*, Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N. Engl. J. Med.* (2004) **350**: 2129-2139.

101. Pao, W., V. Miller, M. Zakowski, J. Doherty, K. Politi, I. Sarkaria, B. Singh, R. Heelan, V. Rusch, L. Fulton, *et al.*, EGF receptor gene mutations are common in lung cancers from "never smokers" and are associated with sensitivity of tumors to gefitinib and erlotinib. *Proc. Natl. Acad. Sci. U.S.A.* (2004) **101**: 13306-13311.

102. Sordella, R., D.W. Bell, D.A. Haber and J. Settleman, Gefitinib-sensitizing EGFR mutations in lung cancer activate anti-apoptotic pathways. *Science* (2004) **305**: 1163-1167.

103. Jiang, J., H. Greulich, P.A. Janne, W.R. Sellers, M. Meyerson and J.D. Griffin, Epidermal growth factor-independent transformation of Ba/F3 cells with cancer-derived epidermal growth factor receptor mutants induces gefitinib-sensitive cell cycle progression. *Cancer Res.* (2005) **65**: 8968-8974.

104. Mitsudomi, T., T. Kosaka and Y. Yatabe, Biological and clinical implications of EGFR mutations in lung cancer. *Int. J. Clin. Oncol.* (2006) **11**: 190-8.

105. Shigematsu, H., T. Takahashi, M. Nomura, K. Majmudar, M. Suzuki, H. Lee, I.I. Wistuba, K.M. Fong, S. Toyooka, N. Shimizu, *et al.*, Somatic mutations of the HER2 kinase domain in lung adenocarcinomas. *Cancer Res.* (2005) **65**: 1642-1646.

106. Sasaki, H., S. Shimizu, K. Endo, M. Takada, M. Kawahara, H. Tanaka, A. Matsumura, K. Iuchi, H. Haneda, E. Suzuki, *et al.*, EGFR and erbB2 mutation status in Japanese lung cancer patients. *Int. J. Cancer* (2006) **118**: 180-4.

107. Buttitta, F., F. Barassi, G. Fresu, L. Felicioni, A. Chella, D. Paolizzi, G. Lattanzio, S. Salvatore, P.P. Camplese, S. Rosini, *et al.*, Mutational analysis of the HER2 gene in lung tumors from Caucasian patients: mutations are mainly present in adenocarcinomas with bronchioloalveolar features. *Int. J. Cancer* (2006) **119**: 2586-91.

108. Lee, J.W., Y.H. Soung, S.Y. Kim, S.W. Nam, W.S. Park, Y.P. Wang, K.H. Jo, S.W. Moon, S.Y. Song, J.Y. Lee, *et al.*, ERBB2 kinase domain mutation in the lung squamous cell carcinoma. *Cancer Lett.* (2006) **237**: 89-94.

109. Lee, J.W., Y.H. Soung, S.H. Seo, S.Y. Kim, C.H. Park, Y.P. Wang, K. Park, S.W. Nam, W.S. Park, S.H. Kim, *et al.*, Somatic mutations of ERBB2 kinase domain in gastric, colorectal, and breast carcinomas. *Clin. Cancer Res.* (2006) **12**: 57-61.

110. Willmore-Payne, C., J.A. Holden and L.J. Layfield, Detection of EGFR- and HER2-activating mutations in squamous cell carcinoma involving the head and neck. *Mod. Pathol.* (2006) **19**: 634-40.

111. Soung, Y.H., J.W. Lee, S.Y. Kim, Y.P. Wang, K.H. Jo, S.W. Moon, W.S. Park, S.W. Nam, J.Y. Lee, N.J. Yoo, *et al.*, Somatic mutations of the ERBB4 kinase domain in human cancers. *Int. J. Cancer* (2006) **118**: 1426-9.

112. Shigematsu, H. and A.F. Gazdar, Somatic mutations of epidermal growth factor receptor signaling pathway in lung cancers. *Int. J. Cancer* (2006) **118**: 257-62.

113. Huang, S.-F., H.-P. Liu, L.-H. Li, Y.-C. Ku, Y.-N. Fu, H.-Y. Tsai, Y.-T. Chen, Y.-F. Lin, W.-C. Chang, H.-P. Kuo, *et al.*, High frequency of epidermal growth factor receptor mutations with complex patterns in non-small cell lung cancers related to gefitinib responsiveness in Taiwan. *Clin. Cancer Res.* (2004) **10**: 8195-8203.

114. Kosaka, T., Y. Yatabe, H. Endoh, H. Kuwano, T. Takahashi and T. Mitsudomi, Mutations of the epidermal growth factor receptor gene in lung cancer: biological and clinical implications. *Cancer Res.* (2004) **64**: 8919-23.

115. Discafani, C.M., M.L. Carroll, M.B. Floyd, Jr., I.J. Hollander, Z. Husain, B.D. Johnson, D. Kitchen, M.K. May, M.S. Malo, A.A. Minnick, Jr., *et al.*, Irreversible inhibition of epidermal growth factor receptor tyrosine kinase with in vivo activity by N-[4-[(3-bromophenyl)amino]-6-quinazolinyl]-2-butynamide (CL-387,785). *Biochem. Pharmacol.* (1999) **57**: 917-25.

116. Sonobe, M., T. Manabe, H. Wada and F. Tanaka, Lung adenocarcinoma harboring mutations in the ERBB2 kinase domain. *J. Mol. Diagn.* (2006) **8**: 351-356.

117. Tam, I.Y., L.P. Chung, W.S. Suen, E. Wang, M.C. Wong, K.K. Ho, W.K. Lam, S.W. Chiu, L. Girard, J.D. Minna, *et al.*, Distinct epidermal growth factor receptor and KRAS mutation patterns in non-small cell lung cancer patients with different tobacco exposure and clinicopathologic features. *Clin. Cancer Res.* (2006) **12**: 1647-53.

118. Bairoch, A., R. Apweiler, C.H. Wu, W.C. Barker, B. Boeckmann, S. Ferro, E. Gasteiger, H. Huang, R. Lopez, M. Magrane, *et al.*, The universal protein resource (UniProt). *Nucleic Acids Res.* (2005) **33**: D154-159.

119. Altschul, S., T. Madden, A. Schaffer, J. Zhang, Z. Zhang, W. Miller and D. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* (1997) **25**: 3389-3402.

120. Edgar, R.C., MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* (2004) **32**: 1792-1797.

121. Mayrose, I., A. Mitchell and T. Pupko, Site-specific evolutionary rate inference: taking phylogenetic uncertainty into account. *J. Mol. Evol.* (2005) **60**: 345-53.

122. Paez, J.G., P.A. Janne, J.C. Lee, S. Tracy, H. Greulich, S. Gabriel, P. Herman, F.J. Kaye, N. Lindeman, T.J. Boggon, *et al.*, EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* (2004) **304**: 1497-1500.

123. Wan, P.T.C., M.J. Garnett, S.M. Roe, S. Lee, D. Niculescu-Duvaz, V.M. Good, C.G. Project, C.M. Jones, C.J. Marshall and C.J. Springer, Mechanism of activation of the RAF-ERK signaling pathway by oncogenic mutations of B-RAF. *Cell* (2004) **116**: 855-867.

124. Kitayama, H., Y. Kanakura, T. Furitsu, T. Tsujimura, K. Oritani, H. Ikeda, H. Sugahara, H. Mitsui, Y. Kanayama and a. Kitamura et, Constitutively activating mutations of c-kit receptor tyrosine kinase confer factor-independent growth and tumorigenicity of factor-dependent hematopoietic cell lines. *Blood* (1995) **85**: 790-798.

125. Mol, C.D., D.R. Dougan, T.R. Schneider, R.J. Skene, M.L. Kraus, D.N. Scheibe, G.P. Snell, H. Zou, B.-C. Sang and K.P. Wilson, Structural basis for the autoinhibition and STI-571 inhibition of c-Kit tyrosine kinase. *J. Biol. Chem.* (2004) **279**: 31655-31663.

126. Glover, H.R., D.A. Baker, A. Celetti and N.J. Dibb, Selection of activating mutations of c-fms in FDC-P1 cells. *Oncogene* (1995) **11**: 1347-1356.

127. Hubbard, S.R., M. Mohammadi and J. Schlessinger, Autoregulatory mechanisms in protein-tyrosine kinases. *J. Biol. Chem.* (1998) **273**: 11987-11990.

128. Cohen, E.E.W., M.W. Lingen, L.E. Martin, P.L. Harris, B.W. Brannigan, S.M. Haserlat, R.A. Okimoto, D.C. Sgroi, S. Dahiya, B. Muir, *et al.*, Response of some head and neck cancers to epidermal growth factor receptor tyrosine kinase inhibitors may be linked to mutation of ERBB2 rather than EGFR. *Clin. Cancer Res.* (2005) **11**: 8105-8108.

129. Slepkov, E.R., J.K. Rainey, B.D. Sykes and L. Fliegel, Structural and functional analysis of the Na+/H+ exchanger. *Biochem. J.* (2007) **401**: 623-633.

130. Orlowski, J. and S. Grinstein, Diversity of the mammalian sodium/proton exchanger SLC9 gene family. *Pflugers Arch.* (2004) **447**: 549-65.

131. Wakabayashi, S., P. Fafournoux, C. Sardet and J. Pouyssegur, The Na+/H+ antiporter cytoplasmic domain mediates growth factor signals and controls "H+-sensing". *Proc. Natl. Acad. Sci. U.S.A.* (1992) **89**: 2424-2428.

132. Shrode, L.D., B.S. Gan, S.J. D'Souza, J. Orlowski and S. Grinstein, Topological analysis of NHE1, the ubiquitous Na+/H+ exchanger using chymotryptic cleavage. *Am. J. Physiol.* (1998) **275**: C431-9.

133. Padan, E., E. Bibi, M. Ito and T.A. Krulwich, Alkaline pH homeostasis in bacteria: new insights. *Biochim. Biophys. Acta* (2005) **1717**: 67-88.

134. Dibrov, P., A. Rimon, J. Dzioba, A. Winogrodzki, Y. Shalitin and E. Padan, 2-Aminoperimidine, a specific inhibitor of bacterial NhaA Na(+)/H(+) antiporters. *FEBS Lett.* (2005) **579**: 373-8.

135. Brett, C.L., M. Donowitz and R. Rao, Evolutionary origins of eukaryotic sodium/proton exchangers. *Am. J. Physiol., Cell Physiol.* (2005) **288**: C223-239.

136. Forrest, L.R., C.L. Tang and B. Honig, On the accuracy of homology modeling and sequence alignment methods applied to membrane proteins. *Biophys. J.* (2006) **91**: 508-17.

137. Wakabayashi, S., T. Pang, X. Su and M. Shigekawa, A novel topology model of the human Na(+)/H(+) exchanger isoform 1. *J. Biol. Chem.* (2000) **275**: 7942-9.

138. Jaroszewski, L., L. Rychlewski, Z. Li, W. Li and A. Godzik, FFAS03: a server for profile-profile sequence alignments. *Nucleic Acids Res.* (2005) **33**: W284-288.

139. Bateman, A., L. Coin, R. Durbin, R.D. Finn, V. Hollich, S. Griffiths-Jones, A. Khanna, M. Marshall, S. Moxon, E.L.L. Sonnhammer, *et al.*, The Pfam protein families database. *Nucleic Acids Res.* (2004) **32**: D138-141.

140. Tang, C.L., L. Xie, I.Y. Koh, S. Posy, E. Alexov and B. Honig, On the role of structural information in remote homology detection and sequence alignment: new methods using hybrid sequence profiles. *J. Mol. Biol.* (2003) **334**: 1043-62.

141. Fleishman, S.J., S.E. Harrington, A. Enosh, D. Halperin, C.G. Tate and N. Ben-Tal, Quasi-symmetry in the cryo-EM structure of EmrE provides the key to modeling its transmembrane domain. *J. Mol. Biol.* (2006) **364**: 54-67.

142. Bowie, J.U., Solving the membrane protein folding problem. *Nature* (2005) **438**: 581-9.

143. von Heijne, G., Membrane-protein topology. *Nat. Rev. Mol. Cell. Biol.* (2006) **7**: 909-18.

144. Murtazina, R., B.J. Booth, B.L. Bullis, D.N. Singh and L. Fliegel, Functional analysis of polar amino-acid residues in membrane associated regions of the NHE1 isoform of the mammalian Na+/H+ exchanger. *Eur. J. Biochem.* (2001) **268**: 4674-85.

145. von Heijne, G., The distribution of positively charged residues in bacterial inner membrane proteins correlates with the trans-membrane topology. *EMBO J.* (1986) **5**: 3021-3027.

146. Ding, J., J.K. Rainey, C. Xu, B.D. Sykes and L. Fliegel, Structural and functional characterization of transmembrane segment VII of the Na+/H+ exchanger isoform 1. *J. Biol. Chem.* (2006) **281**: 29817-29.

147. Counillon, L., J. Noel, R.A. Reithmeier and J. Pouyssegur, Random mutagenesis reveals a novel site involved in inhibitor interaction within the fourth transmembrane segment of the Na+/H+ exchanger-1. *Biochemistry* (1997) **36**: 2951-9.

148. Olami, Y., A. Rimon, Y. Gerchman, A. Rothman and E. Padan, Histidine 225, a residue of the NhaA-Na^+/H^+ antiporter of *Escherichia coli* is exposed and faces the cell exterior. *J. Biol. Chem.* (1997) **272**: 1761-1768.

149. Kaim, G., F. Wehrle, U. Gerike and P. Dimroth, Molecular basis for the coupling ion selectivity of F1F0 ATP synthases: probing the liganding groups for Na+ and Li+ in the c subunit of the ATP synthase from Propionigenium modestum. *Biochemistry* (1997) **36**: 9185-94.

150. Sato, Y. and M. Sakaguchi, Topogenic properties of transmembrane segments of arabidopsis thaliana NHX1 reveal a common topology model of the Na+/H+ exchanger family. *J. Biochem. (Tokyo)* (2005) **138**: 425-431.

151. Galili, L., K. Herz, O. Dym and E. Padan, Unraveling functional and structural interactions between transmembrane domains IV and XI of NhaA Na+/H+ antiporter of Escherichia coli. *J. Biol. Chem.* (2004) **279**: 23104-13.

152. Inoue, H., T. Noumi, T. Tsuchiya and H. Kanazawa, Essential aspartic acid residues, Asp-133, Asp-163 and Asp-164, in the transmembrane helices of a Na+/H+ antiporter (NhaA) from Escherichia coli. *FEBS Lett.* (1995) **363**: 264-8.

153. Kozachkov, L., K. Herz and E. Padan, Functional and structural interactions of the transmembrane domain X of NhaA, Na(+)/H(+) antiporter of escherichia coli, at physiological pH. *Biochemistry* (2007) **46**: 2419-30.

154. Slepkov, E.R., J.K. Rainey, X. Li, Y. Liu, F.J. Cheng, D.A. Lindhout, B.D. Sykes and L. Fliegel, Structural and functional characterization of transmembrane segment IV of the NHE1 isoform of the Na+/H+ exchanger. *J. Biol. Chem.* (2005) **280**: 17863-17872.

155. Rimon, A., Y. Gerchman, Y. Olami, S. Schuldiner and E. Padan, Replacements of histidine 226 of NhaA-Na+/H+ antiporter of Escherichia coli. Cysteine (H226C) or serine (H226S) retain both normal activity and pH sensitivity, aspartate (H226D) shifts the pH profile toward basic pH, and alanine (H226A) inactivates the carrier at all pH values. *J. Biol. Chem.* (1995) **270**: 26813-7.

156. Chang, A.B., R. Lin, W. Keith Studley, C.V. Tran and M.H. Saier, Jr., Phylogeny as a guide to structure and function of membrane transport proteins. *Mol. Membr. Biol.* (2004) **21**: 171-81.

157. Eyre, T.A., F. Ducluzeau, T.P. Sneddon, S. Povey, E.A. Bruford and M.J. Lush, The HUGO gene nomenclature database, 2006 updates. *Nucleic Acids Res.* (2006) **34**: D319-21.

158. Finn, R.D., J. Mistry, B. Schuster-Bockler, S. Griffiths-Jones, V. Hollich, T. Lassmann, S. Moxon, M. Marshall, A. Khanna, R. Durbin, *et al.*, Pfam: clans, web tools and services. *Nucleic Acids Res.* (2006) **34**: D247-251.

159. Eddy, S.R., Hidden Markov models. *Curr. Opin. Struc. Biol.* (1996) **6**: 361-5.

160. Rost, B. and S. O'Donoghue, Sisyphus and prediction of protein structure. *Comput. Appl. Biosci.* (1997) **13**: 345-56.

161. Elofsson, A., A study on protein sequence alignment quality. *Proteins* (2002) **46**: 330-339.

162. Al-Lazikani, B., F.B. Sheinerman and B. Honig, Combining multiple structure and sequence alignments to improve sequence detection and alignment: Application to the SH2 domains of Janus kinases. *Proc. Natl. Acad. Sci. U.S.A.* (2001) **98**: 14796-14801.

163. Jaroszewski, L., L. Rychlewski and A. Godzik, Improving the quality of twilight-zone alignments. *Protein Sci.* (2000) **9**: 1487-96.

164. Khadilkar, A., P. Iannuzzi and J. Orlowski, Identification of sites in the second exomembrane loop and ninth transmembrane helix of the mammalian Na+/H+

exchanger important for drug recognition and cation translocation. *J. Biol. Chem.* (2001) **276**: 43792-800.

165. Counillon, L., A. Franchi and J. Pouyssegur, A point mutation of the Na+/H+ exchanger gene (NHE1) and amplification of the mutated allele confer amiloride resistance upon chronic acidosis. *Proc. Natl. Acad. Sci. U.S.A.* (1993) **90**: 4508-4512.
166. Touret, N., P. Poujeol and L. Counillon, Second-site revertants of a low-sodium-affinity mutant of the Na+/H+ exchanger reveal the participation of TM4 into a highly constrained sodium-binding site. *Biochemistry* (2001) **40**: 5095-5101.
167. Yun, C.H., P.J. Little, S.K. Nath, S.A. Levine, J. Pouyssegur, C.M. Tse and M. Donowitz, Leu143 in the putative fourth membrane spanning domain is critical for amiloride inhibition of an epithelial Na+/H+ exchanger isoform (NHE-2). *Biochem. Biophys. Res. Commun.* (1993) **193**: 532-9.
168. Slepkov, E.R., S. Chow, M.J. Lemieux and L. Fliegel, Proline residues in transmembrane segment IV are critical for activity, expression and targeting of the Na+/H+ exchanger isoform 1. *Biochem. J.* (2004) **379**: 31-8.
169. Lacroix, J., M. Poet, C. Maehrel and L. Counillon, A mechanism for the activation of the Na/H exchanger NHE-1 by cytoplasmic acidification and mitogens. *EMBO rep.* (2004) **5**: 91-6.
170. Fliegel, L., Identification of conserved polar residues important for salt tolerance by the Na+/H+ exchanger of Schizosaccharomyces pombe. *Mol. Cell. Biochem.* (2005) **268**: 83-92.
171. Wiebe, C.A., E.R. Dibattista and L. Fliegel, Functional role of polar amino acid residues in Na+/H+ exchangers. *Biochem. J.* (2001) **357**: 1-10.
172. Wiebe, C.A., C. Rieder, P.G. Young, P. Dibrov and L. Fliegel, Functional analysis of amino acids of the Na+/H+ exchanger that are important for proton translocation. *Mol. Cell. Biochem.* (2003) **254**: 117-124.
173. Noel, J., D. Germain and J. Vadnais, Glutamate 346 of human Na+-H+ exchanger NHE1 is crucial for modulating both the affinity for Na+ and the interaction with amiloride derivatives. *Biochemistry* (2003) **42**: 15361-8.
174. Wang, D., D.F. Balkovetz and D.G. Warnock, Mutational analysis of transmembrane histidines in the amiloride-sensitive Na/H exchanger. *Am. J. Physiol., Cell Physiol.* (1995) **269**: C392-C402.

175. Soong, T.W., T.F. Yong, N. Ramanan and Y. Wang, The Candida albicans antiporter gene CNH1 has a role in Na+ and H+ transport, salt tolerance, and morphogenesis. *Microbiology* (2000) **146**: 1035-44.

176. Simon, E., J. Clotet, F. Calero, J. Ramos and J. Arino, A screening for high copy suppressors of the sit4 hal3 synthetically lethal phenotype reveals a role for the yeast Nha1 antiporter in cell cycle regulation. *J. Biol. Chem.* (2001) **276**: 29740-29747.

177. Watanabe, Y., Y. Shimono, H. Tsuji and Y. Tamai, Role of the glutamic and aspartic residues in Na+-ATPase function in the ZrENA1 gene of Zygosaccharomyces rouxii. *FEMS Microbiol. Lett.* (2002) **209**: 39-43.

178. Mukherjee, S., L. Kallay, C.L. Brett and R. Rao, Mutational analysis of the intramembranous H10 loop of yeast Nhx1 reveals a critical role in ion homoeostasis and vesicle trafficking. *Biochem. J.* (2006) **398**: 97-105.

179. Wang, H., D. Singh and L. Fliegel, Functional role of cysteine residues in the Na+/H+ exchanger effects of mutation of cysteine residues on targeting and activity of the Na+/H+ exchanger. *Arch. Biochem. Biophys.* (1998) **358**: 116-24.

180. Wakabayashi, S., T. Hisamitsu, T. Pang and M. Shigekawa, Mutations of Arg440 and Gly455/Gly456 oppositely change pH sensing of Na+/H+ exchanger 1. *J. Biol. Chem.* (2003) **278**: 11828-35.

181. Counillon, L., J. Pouyssegur and R.A. Reithmeier, The Na+/H+ exchanger NHE-1 possesses N- and O-linked glycosylation restricted to the first N-terminal extracellular domain. *Biochemistry* (1994) **33**: 10463-9.

182. Dibrov, P., P.G. Young and L. Fliegel, Functional analysis of amino acid residues essential for activity in the Na+/H+ exchanger of fission yeast. *Biochemistry* (1998) **37**: 8282-8.

183. Kinclova-Zimmermannova, O., M. Zavrel and H. Sychrova, Identification of conserved prolyl residue important for transport activity and the substrate specificity range of yeast plasma membrane Na+/H+ antiporters. *J. Biol. Chem.* (2005) **280**: 30638-30647.

184. Wakabayashi, S., T. Pang, X. Su and M. Shigekawa, Second mutations rescue point mutant of the Na+/H+ exchanger NHE1 showing defective surface expression. *FEBS Lett.* (2000) **487**: 257-261.

185. Kinclova-Zimmermannova, O., M. Zavrel and H. Sychrova, Importance of the seryl and threonyl residues of the fifth transmembrane domain to the substrate specificity of yeast plasma membrane Na+/H+ antiporters. *Mol. Membr. Biol.* (2006) **23**: 349-61.

APPENDIX I

Nucleic Acids Research, 2005, Vol. 33, Web Server issue ***W299–W302***
doi:10.1093/nar/gki370

ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures

Meytal Landau, Itay Mayrose[1], Yossi Rosenberg, Fabian Glaser[2], Eric Martz[3], Tal Pupko[1] and Nir Ben-Tal*

Department of Biochemistry, George S. Wise Faculty of Life Sciences, Tel Aviv University, Ramat Aviv 69978, Israel, [1]Department of Cell Research and Immunology, George S. Wise Faculty of Life Sciences, Tel Aviv University, Ramat Aviv 69978, Israel, [2]European Bioinformatics Institute, Wellcome Trust Genome Campus, Cambridge, CB10 1SD, UK and [3]Department of Microbiology, University of Massachusetts, Amherst, MA 01003, USA

Received February 5, 2005; Accepted March 3, 2005

ABSTRACT

Key amino acid positions that are important for maintaining the 3D structure of a protein and/or its function(s), e.g. catalytic activity, binding to ligand, DNA or other proteins, are often under strong evolutionary constraints. Thus, the biological importance of a residue often correlates with its level of evolutionary conservation within the protein family. ConSurf (http://consurf.tau.ac.il/) is a web-based tool that automatically calculates evolutionary conservation scores and maps them on protein structures via a user-friendly interface. Structurally and functionally important regions in the protein typically appear as patches of evolutionarily conserved residues that are spatially close to each other. We present here version 3.0 of ConSurf. This new version includes an empirical Bayesian method for scoring conservation, which is more accurate than the maximum-likelihood method that was used in the earlier release. Various additional steps in the calculation can now be controlled by a number of advanced options, thus further improving the accuracy of the calculation. Moreover, ConSurf version 3.0 also includes a measure of confidence for the inferred amino acid conservation scores.

INTRODUCTION

The degree to which an amino acid position is recessive to substitutions is strongly dependent on its structural and functional importance. An amino acid that plays an essential role, e.g. in enzymatic catalysis, is likely to remain unaltered in spite of the random evolutionary drift. Hence, the level of evolutionary conservation is often indicative of the importance of the position in maintaining the protein's structure and/or function.

ConSurf is a web server for mapping the level of evolutionary conservation of each of the amino acid positions of a protein onto its 3D structure (1). The conservation scores are calculated based on the evolutionary relations among the protein and its homologs and the probability of residue replacement as reflected in amino acid substitution matrices (2,3). The scores are subsequently translated into a discrete coloring scale that is used to project them on a known 3D structure of one of the homologous proteins. The server is implemented in a user-friendly interface that enables scientists from the experimental biology as well as the bioinformatics communities to explore the evolutionary history of a protein of known 3D structure and to identify structurally and functionally important positions. We provide here a brief review of ConSurf with emphasis on the new features that were added recently.

METHODS

A short description of the methodology is provided here and a more detailed description is available at http://consurf.tau.ac.il/, under 'OVERVIEW', 'QUICK HELP' and 'FAQ'.

ConSurf protocol

A flow chart, describing the ConSurf protocol, is presented in Figure 1. The minimal input requirement for ConSurf is a four-letter PDB (4) code and the relevant chain identifier of the query protein. Alternatively, a user-provided protein structure in the form of a PDB file can be uploaded. Using the 3D structure of the query protein as an input, the following steps are automatically carried out by ConSurf:

(i) The amino acid sequence is extracted from the PDB file.
(ii) Homologous sequences in the SWISS-PROT database (5) are searched and collected using PSI-BLAST (6).
(iii) A multiple sequence alignment (MSA) of these sequences is computed using CLUSTAL W (7).

*To whom correspondence should be addressed. Tel: +972 3 640 6709; Fax: +972 3 640 6834; Email: bental@ashtoret.tau.ac.il

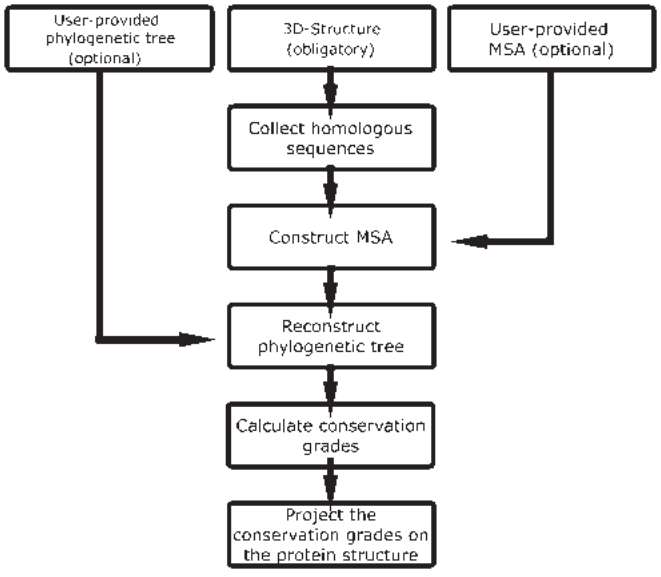


Figure 1. A flow chart of a ConSurf calculation.

(iv) A phylogenetic tree is reconstructed based on the MSA, using the neighbor-joining algorithm (8) as implemented in the Rate4Site program (3).

(v) Position-specific conservation scores are computed using the empirical Bayesian (2) or maximum-likelihood (3) algorithms.

(vi) The continuous conservation scores are divided into a discrete scale of 9 grades for visualization purpose. Grade 1 contains the most variable positions and is colored turquoise; grade 5 contains intermediately conserved positions and is colored white; and grade 9 contains the most conserved positions and is colored maroon.

(vii) The nine-color conservation grades are projected onto the 3D structure of the query protein.

The sensitivity and selectivity of the search for homologous proteins [step (ii) above] can be controlled by adjusting the number of PSI-BLAST iterations, the PSI-BLAST *E*-value cut-off and the maximum number of sequences extracted from PSI-BLAST (6). As an alternative to this automatic search, the server accepts a user-provided MSA. In such a case, steps (ii) and (iii) in the outline protocol are skipped.

ConSurf outputs

After the calculation begins, ConSurf produces a status page indicating the computation parameters along with the different stages of the server activity. The main result of a ConSurf calculation is under the link 'View ConSurf Results with Protein Explorer', which leads to the graphic visualization of the query protein, color coded by conservation scores, through the Protein Explorer interface (9). The continuous conservation scores of each of the amino acid positions are available under the link 'Amino Acid Conservation Scores', along with the color grades and additional data. The script command for viewing the 3D structure of the query protein, color coded by conservation scores, is available under the link 'RasMol coloring script source'. This file can be downloaded and used locally with the RasMol program (10), thus producing the same color-coded scheme generated by the server. A PDB file, in which the conservation scores are specified in the temperature (B) factor field, can be downloaded through the link: 'The PDB file updated with the conservation scores in the tempFactor field'. Thus, any 3D protein viewer, such as the RasMol program (10), which is capable of presenting the B factors, is suitable for mapping the conservation scores on the structure.

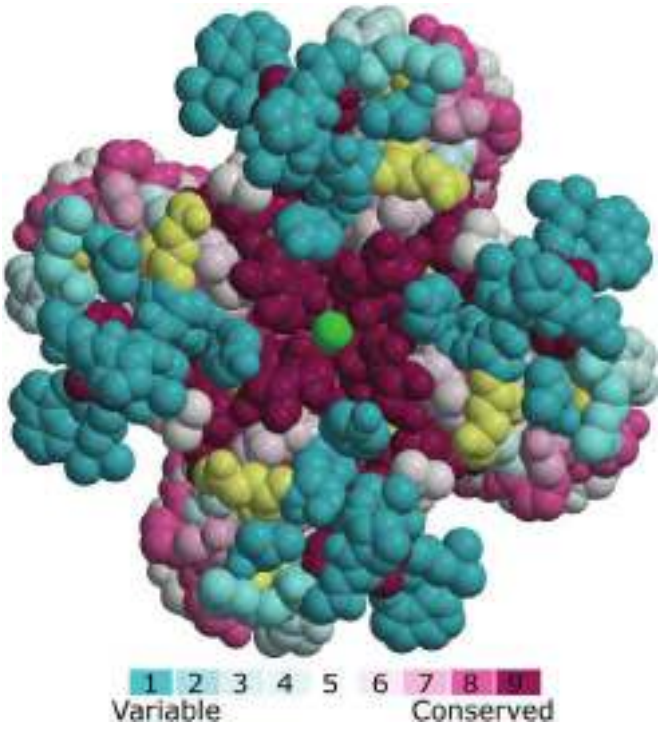


Figure 2. A ConSurf analysis of the Kcsa potassium channel. The tetrameric channel, which is viewed along the pore from the extracellular end, is presented using a space-filled model. The amino acids are colored by their conservation grades using the color-coding bar, with turquoise-through-maroon indicating variable-through-conserved. Amino acid positions, for which the inferred conservation level was assigned with low confidence, are marked with light yellow. The potassium ion at the channel pore is colored green. Conservation scores, which were calculated for one of the channel's subunits, were projected on the homotetrameric structure. The run was carried out using PDB code 1bl8 (11) and default ConSurf parameters. The picture was generated using MOLSCRIPT (21) and Raster3D (26).

The ConSurf output also includes links to the PSI-BLAST results, the homologous sequences along with a link to their SWISS-PROT entry page, the MSA and the phylogenetic tree used in the calculation.

As an example, we provide in Figure 2 the main output of a ConSurf run of the Kcsa potassium-channel (11), a transmembrane protein from *Streptomyces Lividans*. Kcsa is a homotetramer with a 4-fold symmetry axis about its pore. The ConSurf calculations demonstrate the high level of conservation of the pore region as compared with the rest of the protein. The pore architecture provides the unique stereochemistry which is required for efficient and selective conduction of potassium ions (11). The biological importance of this stereochemistry is reflected by a strong evolutionary pressure to resist amino acid replacements in the pore. In contrast, the regions that surround the pore and face the extracellular matrix are highly variable.

NEW ADDITIONS AND IMPROVEMENTS IN ConSurf

An empirical Bayesian method to score conservation

The heart of the ConSurf server is the calculation of the conservation scores of each amino acid position. In the previous version of the server, the maximum-likelihood method (3) was used as the default option to this end. Recently, we showed that an empirical Bayesian method can significantly improve the accuracy of the estimated conservation scores (2). The empirical Bayesian method is particularly superior to the maximum-likelihood method when the number of homologous sequences analyzed is small (2). The new method is now integrated in ConSurf as the default option. The usage of the maximum-likelihood method is still available under the 'Method' pull-down menu.

Estimation of the reliability of the inferred conservation scores

An amino acid position that is conserved across all homologous sequences will always be assigned with the highest conservation grade. Yet, there is a difference if the conservation score is inferred based on a small MSA of, for example, 4 sequences, or based on a larger set of 30 sequences. Additionally, since the conservation calculation for positions with a lot of gaps is based on a fewer number of sequences, the conservation score for these positions will be less reliable than positions that have no gaps. The reliability of the conservation computation is not only determined by the number of sequences in each position but also by the evolutionary distances between the sequences, the phylogenetic tree topology and the evolutionary process.

One of the most important new features in ConSurf version 3.0 is the inclusion of a measure of the confidence of each of the inferred position-specific conservation scores. The measure is calculated using the empirical Bayesian method, as explained in (12) and at http://consurf.tau.ac.il/ under 'OVERVIEW'. In short, it is based on a confidence interval that is defined by the lower and upper quartiles: the 25th and 75th percentiles of the inferred distribution of conservation scores, respectively. It gives the 50% confidence interval and also indicates the dispersion of each of the estimated scores. The confidence interval is usually large in positions with a small number of sequences, thus indicating a low level of support in the inferred conservation scores for these positions. When the number of sequences is large, the confidence interval is small, and the point score estimates are more assured. Amino acid positions, associated with confidence intervals that are too large to be trustworthy, are marked in the output files of the server and highlighted (in pale yellow) on the 3D structure of the protein (Figure 2).

Models of amino acid substitutions

The inference of the evolutionary conservation scores relies on a specified probabilistic model of amino acid replacements (3). The JTT matrix (13) was used to this end in the previous version of ConSurf. In version 3.0, we expanded the utility of ConSurf to support additional models of substitution for nuclear DNA-encoded as well as non-nuclear DNA-encoded proteins. The model of substitution can be chosen from the 'Model of substitution for proteins' pull-down menu, which is available in the 'Advance Options' section of the ConSurf main interface. The JTT (13), Dayhoff (14) and WAG (15) matrices are suitable for nuclear DNA-encoded proteins. The WAG matrix has been inferred from a large database of sequences comprising a broad range of protein families, and is thus suitable for distantly related amino acid sequences (15). The mtREV (16) and cpREV (17) matrices are suitable for mitochondrial and chloroplast DNA-encoded proteins, respectively. Examples that demonstrate the influence of using the different matrices on the calculations are available at http://consurf.tau.ac.il/ under 'OVERVIEW'. The differences between ConSurf calculations using different matrices tend to be small but not negligible.

User-provided phylogenetic tree

A user-provided phylogenetic-tree (that should be consistent with the MSA) may be supplied as an additional input. In this case, steps (ii–iv) in the 'ConSurf protocol' (specified above) are skipped. We note that the accuracy of the conservation scores calculations relies on the correct reconstruction of the phylogeny (18). Default ConSurf runs are carried out using phylogenetic trees that are constructed with the neighbor-joining algorithm. The new feature enables the users to supply more accurate trees.

WORK UNDER DEVELOPMENT

We are currently integrating a few more enhancements to ConSurf. At present, ConSurf uses the neighbor-joining algorithm as a fast heuristic method to construct phylogenetic trees. Notwithstanding, the more exhaustive maximum-likelihood tree-reconstruction method is known to produce more accurate phylogenetic trees (19), which should increase the accuracy of the calculated conservation scores (18). We will integrate the maximum-likelihood-based SEMPHY program (20) into ConSurf. This program reconstructs phylogenetic trees dramatically faster than other maximum-likelihood tree-reconstruction methods (20), and can thus be used with little additional computational cost.

A computational tool will be developed, which will enable a simultaneous online view of the phylogenetic tree while analyzing the evolutionary profile of the protein. This interactive tool will allow the user to mark specific branches, which will be used for in-depth ConSurf analyses. For example, the selection of specific clades (sub-trees) may be used to define sub-families. The examination of ConSurf analysis of sub-families may reveal specific characters that are unique to each of them.

The main output of ConSurf is the projection of the conservation scores on the 3D structure of the query protein. In order to easily generate high-resolution color figures, we will provide a script command for the MOLSCRIPT program (21) as an additional output.

A planned enhancement to ConSurf will be the inclusion of all the visualization results in the header of the PDB file. The format that will be used to this end will allow an interactive offline view of the results using Protein Explorer on the user machine, exactly as they appear online.

CONCLUSIONS

ConSurf (1) is a web server that automatically calculates evolutionary conservation scores for each amino acid position and projects them onto the 3D structure of the protein. Evolutionary trace (ET) (22,23) is a related web server that may also be used to map conservation scores on the 3D structure. However, the ET method (23), which was developed for the identification of class-specific residues, is less accurate than ConSurf for scoring conservation (3). This may explain why biologically important regions that were detected using ConSurf were overlooked by the ET web server (1). (See http://consurf.tau.ac.il/, under 'OVERVIEW' for details). Other related web servers, such as MSA3D (9), ProtSkin (24) and COLORADO3D (25), may also be used to present conservation scores on protein structures. These web servers use a consensus approach to infer conservation, which is inferior to methods, such as the ET and ConSurf's maximum-likelihood and empirical Bayesian that explicitly take into account the phylogeny of the homologous sequences under study (2,3). Moreover, all the above servers are not fully automated as ConSurf and require a user-provided MSA.

The new version of ConSurf includes an improved algorithm for scoring evolutionary conservation and provides an index of confidence in the estimated scores. In addition, while ConSurf is still easy to use with default options, expert users can benefit from several advanced options that were added in order to provide more control over the calculations and so to increase the accuracy of the results.

ACKNOWLEDGEMENTS

The authors are grateful to the Bioinformatics Unit and the George S. Wise Faculty of Life Sciences at Tel Aviv University for providing technical assistance and computation facilities. This study was supported by a Research Career Development Award from the Israel Cancer Research Fund. T.P. was supported by a grant in Complexity Science from the Yeshaia Horvitz Association and from an ISF grant no. 1208/04. Development of Protein Explorer is supported by a grant to E.M. from the Division of Undergraduate Education of the US National Science Foundation. Funding to pay the Open Access publication charges for this article was provided by Tel Aviv University.

Conflict of interest statement. None declared.

REFERENCES

1. Glaser,F., Pupko,T., Paz,I., Bell,R.E., Bechor-Shental,D., Martz,E. and Ben-Tal,N. (2003) ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. *Bioinformatics*, **19**, 163–164.
2. Mayrose,I., Graur,D., Ben-Tal,N. and Pupko,T. (2004) Comparison of site-specific rate-inference methods for protein sequences: empirical Bayesian methods are superior. *Mol. Biol. Evol.*, **21**, 1781–1791.
3. Pupko,T., Bell,R.E., Mayrose,I., Glaser,F. and Ben-Tal,N. (2002) Rate4Site: an algorithmic tool for the identification of functional regions in proteins by surface mapping of evolutionary determinants within their homologues. *Bioinformatics*, **18**, S71–S77.
4. Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.
5. Bairoch,A. and Apweiler,R. (1999) The SWISS-PROT protein sequence data bank and its supplement TrEMBL in 1999. *Nucleic Acids Res.*, **27**, 49–54.
6. Altschul,S., Madden,T., Schaffer,A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
7. Thompson,J.D., Higgins,D.G. and Gibson,T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **22**, 4673–4680.
8. Saitou,N. and Nei,M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.*, **4**, 406–425.
9. Martz,E. (2002) Protein Explorer: easy yet powerful macromolecular visualization. *Trends Biochem. Sci.*, **27**, 107–109.
10. Sayle,R.A. and Milner-White,E.J. (1995) RASMOL: biomolecular graphics for all. *Trends Biochem. Sci.*, **20**, 374.
11. Doyle,D.A., Morais Cabral,J., Pfuetzner,R.A., Kuo,A., Gulbis,J.M., Cohen,S.L., Chait,B.T. and MacKinnon,R. (1998) The structure of the potassium channel: molecular basis of K+ conduction and selectivity. *Science*, **280**, 69–77.
12. Susko,E., Inagaki,Y., Field,C., Holder,M.E. and Roger,A.J. (2002) Testing for differences in rates-across-sites distributions in phylogenetic subtrees. *Mol. Biol. Evol.*, **19**, 1514–1523.
13. Jones,D.T., Taylor,W.R. and Thornton,J.M. (1992) The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.*, **8**, 275–282.
14. Dayhoff,M.O., Hunt,L.T., Barker,W.C., Schwartz,R.M. and Orcutt,B.C. (1978) In Young,C.L. (eds) *Atlas of Protein Sequence and Structure*. National Biomedical Research Foundation, Washington, DC.
15. Whelan,S. and Goldman,N. (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol. Biol. Evol.*, **18**, 691–699.
16. Adachi,J. and Hasegawa,M. (1996) Model of amino acid substitution in proteins encoded by mitochondrial DNA. *J. Mol. Evol.*, **42**, 459–468.
17. Adachi,J., Waddell,P.J., Martin,W. and Hasegawa,M. (2000) Plastid genome phylogeny and a model of amino acid substitution for proteins encoded by chloroplast DNA. *J. Mol. Evol.*, **50**, 348–358.
18. Mayrose,I., Mitchell,A. and Pupko,T. (2005) Site-specific evolutionary rate inference: taking phylogenetic uncertainty into account. *J. Mol. Evol.*, in press.
19. Felsenstein,J. (1996) Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods. *Methods Enzymol.*, **266**, 418–427.
20. Friedman,N., Ninio,M., Pe'er,I. and Pupko,T. (2002) A structural EM algorithm for phylogenetic inference. *J. Comput. Biol.*, **9**, 331–353.
21. Kraulis,P.J. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.*, **24**, 946–950.
22. Innis,C.A., Shi,J. and Blundell,T.L. (2000) Evolutionary trace analysis of TGF-beta and related growth factors: implications for site-directed mutagenesis. *Protein Eng.*, **13**, 839–847.
23. Lichtarge,O., Bourne,H.R. and Cohen,F.E. (1996) An evolutionary trace method defines binding surfaces common to protein families. *J. Mol. Biol.*, **257**, 342–358.
24. Deprez,C., Lloubes,R., Gavioli,M., Marion,D., Guerlesquin,F. and Blanchard,L. (2005) Solution structure of the *E.coli* TolA C-terminal domain reveals conformational changes upon binding to the phage g3p N-terminal domain. *J. Mol. Biol.*, **346**, 1047–1057.
25. Sasin,J.M. and Bujnicki,J.M. (2004) COLORADO3D, a web server for the visual analysis of protein structures. *Nucleic Acids Res.*, **32**, W586–W589.
26. Merritt,E.A. and Bacon,D.J. (1997) Raster3D photorealistic molecular graphics. *Methods Enzymol.*, **277**, 505–524.

APPENDIX II

Structure, Vol. 12, 2265–2275, December, 2004, ©2004 Elsevier Ltd. All rights reserved. DOI 10.1016/j.str.2004.10.006

A Putative Mechanism for Downregulation of the Catalytic Activity of the EGF Receptor via Direct Contact between Its Kinase and C-Terminal Domains

Meytal Landau, Sarel J. Fleishman, and Nir Ben-Tal*
Department of Biochemistry
George S. Wise Faculty of Life Sciences
Tel-Aviv University
Ramat-Aviv 69978
Israel

Summary

Tyrosine kinase receptors of the EGFR family play a significant role in vital cellular processes and in various cancers. EGFR members are unique among kinases, as the regulatory elements of their kinase domains are constitutively ready for catalysis. Nevertheless, the receptors are not constantly active. This apparent paradox has prompted us to seek mechanisms of regulation in EGFR's cytoplasmic domain that do not involve conformational changes of the kinase domain. Our computational analyses, based on the three-dimensional structure of EGFR's kinase domain suggest that direct contact between the kinase and a segment from the C-terminal regulatory domains inhibits enzymatic activity. EGFR activation would then involve temporal dissociation of this stable complex, for example, via ligand-induced contact formation between the extracellular domains, leading to the reorientation of the transmembrane and intracellular domains. The model provides an explanation at the molecular level for the effects of several cancer-causing EGFR mutations.

Introduction

The epidermal growth factor receptor (EGFR) family of receptor tyrosine kinases (RTKs), also known as ErbB or HER, consists of four members, ErbB1, -2, -3, and -4 (Schlessinger, 2000). The receptors, which are activated by some dozen ligands, including EGF and TGFα, play an important role in the control of many fundamental cellular processes (Schlessinger, 2000). Mutations and overexpression of the ErbBs have been implicated in malignant diseases such as carcinoma and glioblastoma and are linked with aggressive disease, resistance to chemotherapy, and poor survival (Dancey, 2004). Accordingly, the ErbBs are attractive targets for anticancer drugs (Cho et al., 2003; Yarden and Sliwkowski, 2001). Structurally, the ErbBs consist of an N-terminal, extracellular domain that is connected by a short transmembrane span to a tyrosine kinase domain, which is in turn followed by a C-terminal domain.

In all RTKs, including the ErbBs, the active kinase triggers a wide spectrum of crucial intracellular signaling events (Schlessinger, 2000), and their catalytic activity is encapsulated in multiple layers of regulation (Huse and Kuriyan, 2002). A primary means of regulation in RTKs is ligand binding to the extracellular domain, leading to dimerization or formation of higher-order oligomers of the receptors and enzymatic activation (Schlessinger, 2000, 2003). Similarly, activation of ErbB1, -3, and -4 involves ligand-induced contact formation between the extracellular domains of different members of the ErbB family to form homo- and heterodimers (Schlessinger, 2000). Some studies have shown that, without a ligand, EGFR exists mostly in a monomeric form and that ligands induce its dimerization and activation (Yarden and Schlessinger, 1987). On the other hand, recent studies have demonstrated that while required, dimerization is not sufficient for activation and that in the absence of a ligand, stable, inactive dimers exist in a form in which contact between monomers involves the transmembrane and intracellular domains (Biswas et al., 1985; Gadella and Jovin, 1995; Moriki et al., 2001; Yu et al., 2002). Experimental evidence (Cadena et al., 1994), as well as the computational results presented below, demonstrates that the C-terminal domain plays a role in such contact formation.

In most tyrosine kinases (TKs) excluding the ErbBs, an important means of regulation involves profound structural changes along with transautophosphorylation of the kinase domain (Schlessinger, 2000). In contrast, the ErbB family is unique in that activation is independent of its phosphorylation state (Gotoh et al., 1992). The structure of the apo-EGFR kinase domain demonstrated that its unphosphorylated conformation was, in essence, identical to the phosphorylated conformations of other TKs (Stamos et al., 2002).

Recently, a structure of the kinase domain of the EGFR in complex with the inhibitor GW572016 (Lapatinib) was determined (Wood et al., 2004). This structure shows several differences, including different conformations of the substrate and ATP binding sites (Wood et al., 2004), from either the structure of the apo-EGFR or of EGFR bound to the OSI-774 (Tarceva) inhibitor (Stamos et al., 2002). The authors have suggested that these differences are due to the fact that the conformation seen in the GW572016 bound kinase domain reflects an inactive state that is accessible to the kinase domain under physiological conditions. However, GW572016 is very bulky in comparison to OSI-774. Thus, as the authors indicated, another possibility is that the differences in the structures are due to the inhibitor's large size, which forces a conformation that is far from native. That the apo-EGFR kinase domain is seemingly in a constitutively active conformation (Stamos et al., 2002) leads to an apparent paradox, since it is well established that ErbBs are not constitutively active (Schlessinger, 2000). Hence, our working hypothesis, as presented in Figure 1, was that ErbBs are regulated by another mechanism intrinsic to the intracellular domain; one that is phosphorylation independent.

The orphan receptor ErbB2 presents an even more intriguing case than other members of the EGFR family because its activation is not only phosphorylation inde-

*Correspondence: bental@ashtoret.tau.ac.il

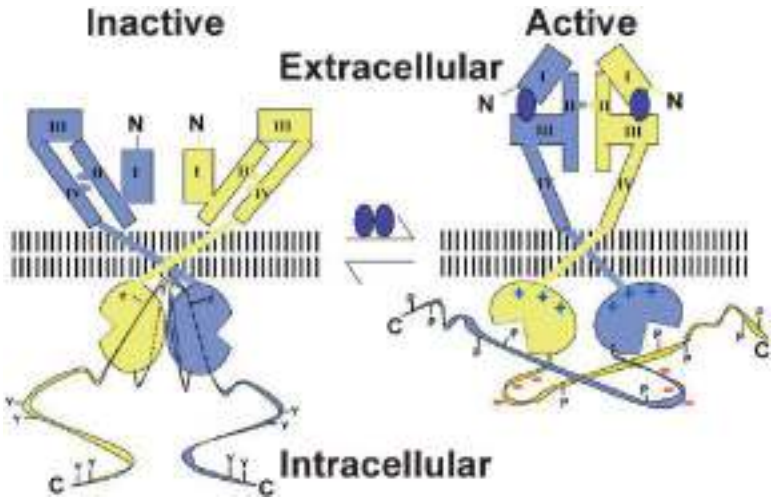


Figure 1. Schematic Diagram Representing the Suggested Model of EGFR Activation

Two EGFR monomers are colored light purple and yellow. The extracellular domain (residues 1–620, labeled I, II, III, and IV according to its subdomains) and the kinase domain (residues 685–957) are connected via a transmembrane helix (residues 621–642) and a short juxtamembrane segment (not shown). The C-terminal domain, comprising 229 amino acids, whose structure has not been determined, follows the kinase domain. Tyrosine residues (Y) known as the autophosphorylation sites in the C-terminal domain are indicated. In the inactive conformation (left), each of the extracellular domains assumes a compact structure, and the intracellular domains contact via the C-terminal fragments, leading to an inactive and stable form. Activation (right) occurs when ligands (purple ovals) bind to the extracellular domains, leading to the formation of a stable extracellular contact, which is followed by the rotation of the transmembrane helices and the subsequent destabilization of the contacts between the C-terminal and kinase domains. The kinase can now transautophosphorylate the tyrosine residues of its own C-terminal domain, as well as tyrosine residues of its protein substrates. The figure displays an illustration of the transmembrane domain; the suggested molecular model for the transmembrane domain in the active and inactive states was presented in (Fleishman et al., 2002). Positive and negative charges are marked in the active conformation on the kinase and C-terminal domains, respectively. In the inactive conformation, they roughly neutralize each other (Figure 2).

pendent, but also ligand independent (Cho et al., 2003). The absence of clear regulation of ErbB2 activation prompted us to propose a molecular mechanism for rotation-coupled activation of this receptor (Fleishman et al., 2002). Specifically, the transmembrane domain of an ErbB2 homodimer may occupy one of two stable conformations, corresponding to the active and inactive states of the receptor. The switch between the two conformations, involving a rotation of the transmembrane domain (Jiang and Hunter, 1999), induces the reorientation of the cytoplasmic domains within receptor dimers, thus leading to transautophosphorylation and stimulation of enzymatic activity. In this paper, we shall analyze the implications of this mode of activation on the conformation of the intracellular kinase domain.

The C-terminal domain plays a role in the internalization and degradation of the EGFR (Chang et al., 1995) and in EGFR's regulation by other molecules (Huse and Kuriyan, 2002). This domain also serves as a docking site for protein modules that bind the phosphotyrosines on the activated receptors (Schlessinger, 2000). In addition to these roles, the importance of the C-terminal domain for proper functioning of the EGFR was previously noted on the basis of studies of viral and other mutant EGFR members (Boerner et al., 2003; Chang et al., 1995; Wedegaertner and Gill, 1992).

Naturally occurring retroviral oncogene variants (v-ErbB) are an extracellular domain-truncated form of the EGFR gene that affects cell growth, motility, and survival (Gamett et al., 1986). These v-ErbB variants share striking homology with mutants of the human EGFR members that have been identified in gliomas and carcinomas (Frederick et al., 2000). Truncation of the extracellular domain is insufficient to manifest the transforming properties of the different v-ErbB variants; these properties are probably related to amino acid replacements, insertions, and deletions in the C-terminal domain (Boerner et al., 2003; Massoglia et al., 1990; Riedel et al., 1987). Variations in the C-terminal domain of ErbB receptors are known to be responsible for the alterations in the transforming potential and type of malignant diseases due to the expression of v-ErbBs in affected cells (Gamett et al., 1986; Pelley et al., 1989; Raines et al., 1988). The increased substrate-phosphorylation capacity of the C-terminally impaired EGFR is not attributed to lesser degradation and internalization, but rather to an enhanced rate of autophosphorylation (Robinson et al., 1992), thus providing direct evidence for a relationship between C-terminal domain impairment and increased catalysis.

Here, we propose a molecular model clarifying some of the ambiguity regarding the role of the C-terminal domain in ErbB regulation. According to the model (Figure 1), contact between the intracellular domains of the EGFR within a dimer leads to receptor inactivation, while ligand-induced contact between the extracellular domains leads to rotation-coupled activation (Fleishman et al., 2002; Jiang and Hunter, 1999; Moriki et al., 2001) by destabilization of the intermonomer contacts in the cytoplasmic domain. According to this scenario, interactions between the intracellular domains regulate activation (Burgess et al., 2003; Chantry, 1995), and the C-terminal domain acts as an inherent negative regulator of the EGFR's activity. This model offers a molecular mechanism that underlies the tumorigenic activity of EGFR mutants.

Results

Geometric Complementarity between the Kinase and C-Terminal Domains

The crystal structure of the EGFR (Stamos et al., 2002) (PDB entries 1m14 and 1m17) includes the kinase domain (residues 685–957) and a segment from the

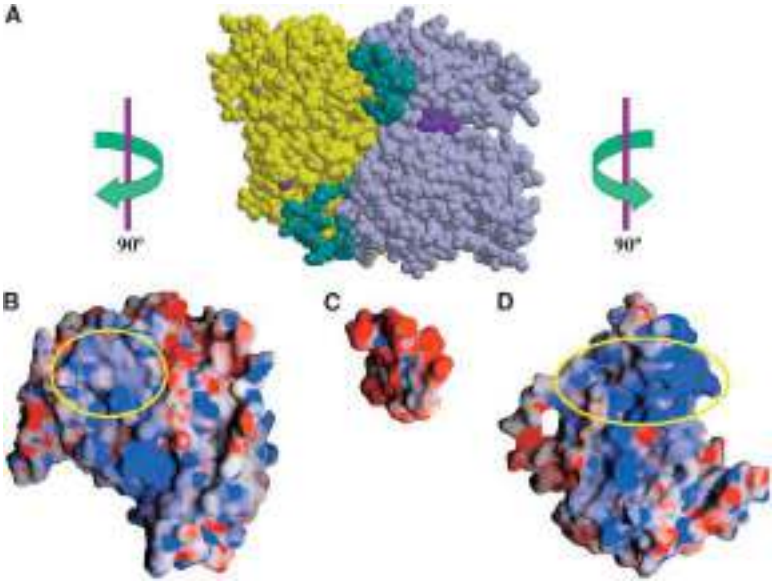


Figure 2. Geometric and Electrostatic Complementarity between the Kinase and C-Terminal Domains

(A) A space-filled model of EGFR's homodimeric complex (Stamos et al., 2002) showing the geometric complementarity within the complex. The kinase domains are colored light purple and yellow, the C-terminal fragments are colored cyan, and the inhibitor is colored purple. The dimer is symmetric, which means that each kinase domain is in contact with both C-terminal fragments, yielding one large and one small interface per monomer. The interactions with the C terminus are identical in both monomers. Figure 2A was made by using MOLSCRIPT (Kraulis, 1991) and Raster3D (Merritt and Bacon, 1997).

(B–D) A projection of the electrostatic potential (ϕ) onto the molecular surface of the kinase domain and the C-terminal fragment that comprise the complex in (A); $\phi > 10$ kT/e is dark blue, $\phi = 0$ is white, and $\phi < -10$ kT/e is dark red. Yellow ellipses mark the interfaces between the kinase domains and the C-terminal fragments. The figures were produced by using GRASP (Nicholls et al., 1991). (B) The left-most kinase domain shown in (A) (yellow) was rotated 90° to the left relative to its orientation in (A). (C) The C-terminal fragment is shown in the same orientation as the upper segment in (A). (D) The right-most kinase domain shown in (A) (light purple) was rotated 90° to the right relative to its orientation in (A). The electrostatic complementarity between the negatively charged C-terminal fragment and the positively charged residues of the kinase domain that interact with it is noticeable.

C-terminal domain (residues 977–995). The crystal structure reveals six putative dimer forms (Stamos et al., 2002). We focus here on the one with the largest intersubunit interface. In this complex, the kinase domain was found as a symmetric homodimer (Figure 2A), in which two copies of the fragment of the C-terminal domain mediate contact between the two kinase domains. This dimer is also the only one in which the kinase domains' N termini are facing in the same direction, in accordance with the physiological requirement that the two domains connect to the membrane bilayer.

We calculated the water-accessible surface area of the kinase domain alone and within the homodimeric complex. Each kinase monomer contacts two C-terminal fragments (Figure 2A). The water-accessible surface areas of these interfaces are 1419 Å^2 and 1048 Å^2. Thus, the total interface between each monomer of the kinase domain and the C-terminal fragments is 2467 Å^2, and the interface within the entire complex is twice as large, constituting a very large interface compared to typical interprotein interfaces (Jones and Thornton, 1996).

Charge Complementarity between the Kinase and C-Terminal Domains

Electrostatic calculations show strong positive potential in the kinase domain at its interface with the C-terminal fragments (Figures 2B and 2D). This potential originates from positively charged residues in both subunits, suggesting that the kinase domains would repel one another in the absence of the C-terminal fragments. Kinase domains from other ErbBs, which were constructed by using homology modeling, displayed similar positive electrostatic potentials in the corresponding regions (data not shown). The C-terminal fragment of all of the ErbBs contained 8–10 acidic and no basic residues (Figure 3). These residues produced a highly negative electrostatic potential (Figure 2C). Thus, the kinase domain and the C-terminal fragments form complementary surfaces in terms of their electrostatic potential. The geometric and charge complementarity (Figures 2B–2D), together with the significant size of the interface (Figure 2A), are indicative of the stability of the complex and suggest that it may be biologically meaningful.

Following the experiments of Chang et al. (1995) discussed below, we simultaneously substituted each of the negatively charged residues 979–982 (DEED, Figure 3) in the C-terminal domain with its polar equivalent, i.e., D→N and E→Q. The mutated C-terminal fragment is much less negatively charged than the native fragment (Figure 4B), and this difference in charge obstructs its electrostatic complementarity with the kinase domain and presumably destabilizes the complex. We further mutated the same positions to four positively charged lysine residues (Figure 4C). Electrostatic analysis of the mutated C-terminal fragment displayed a positive potential at the N-terminal region of the fragment, which would lead to its electrostatic repulsion from the kinase domain. To test whether the charge complementarity is

Figure 3. Abundance of Acidic Residues in the Fragment of the C-Terminal Domain

The multiple sequence alignment of the C-terminal segments of the four human members of the ErbB family. Each segment contains between 8 and 10 acidic residues (marked in red).

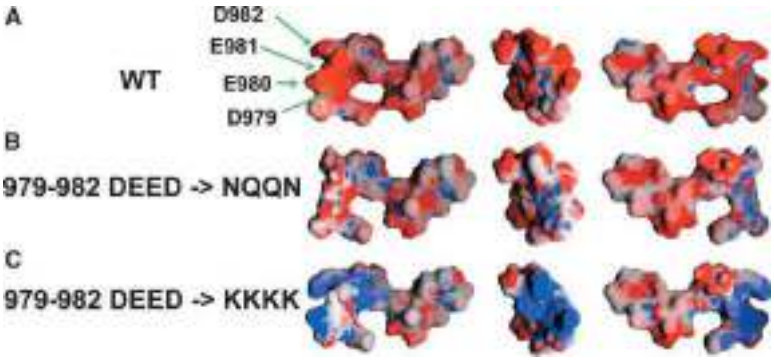


Figure 4. Electrostatic Analysis of Reported and Novel Mutations in the EGFR C-Terminal Fragment

A projection of the electrostatic potential (ϕ) onto the molecular surface of the C-terminal fragment; the color coding is as in Figures 2B–2D.

(A) The C-terminal fragment of the native EGFR, in the same orientation as in Figure 2C (central image), rotated 90° to the right (right image) or left (left image). The location of selected residues is marked.

(B) The C-terminal fragment, in the same orientations as in (A), in which the negatively charged residues in the 979–982 positions (DEED) were mutated to their polar equivalent (NQQN).

(C) The C-terminal fragment, in the same orientations as in (A), in which the same positions were mutated to positively charged lysine residues.

unique to ErbBs among TKs, we examined the electrostatic potentials of a few TKs of known structures as described in the Supplemental Data (available with this article online; Electrostatic calculations). These domains, which were derived from remotely related proteins, display diverse electrostatic characteristics. In particular, they do not share EGFR's strong positive electrostatic potential at the interface with the C-terminal fragments (data not shown), suggesting that such electrostatic interactions between the kinase and the C-terminal domains are specific to the ErbBs.

A Network of Ion Pairs and Hydrogen Bonds at the Interface

Our analysis demonstrated that a network of salt bridges and hydrogen bonds connects the two adjacent kinase domains through the C-terminal fragments (Figure 5B). We identified four charged residues within this network that are involved in several interactions with neighboring residues and are buried at the interface of the EGFR complex. Of these residues, two are positively charged (Lys822 and Lys828 on the kinase domain) and two are negatively charged (Asp988 and Asp990 on the C-terminal fragment) (Figure 5B).

Polar networks, such as the one observed in the EGFR interface (Figure 5B), significantly increase the stability of complexes and contribute to the binding specificity (Sheinerman et al., 2000). Therefore, mutations of charged positions in the network would alter the stability of the complex (Serrano et al., 1990). An even larger effect would be obtained by mutating them in pairs. For example, mutating Lys822 and Lys828 to aspartates or Asp988 and Asp990 to lysines altered the electrostatic surface of the kinase domain and the C-terminal fragment, respectively (Figure 6). Such mutations would impinge on the formation of the EGFR complex and kinase activation.

The importance of the network for the stability of the

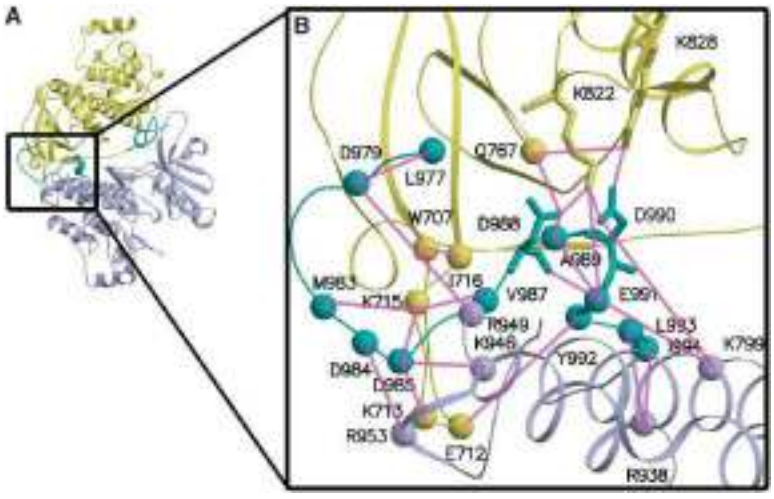


Figure 5. A Network of Ion Pairs and Hydrogen Bonds across the Interface of the EGFR Complex

The kinase domain monomers are displayed as ribbons and colored light purple and yellow. The C-terminal fragment is colored cyan.

(A) The EGFR homodimeric complex (Stamos et al., 2002) as viewed with a clockwise rotation of about 90° compared to Figure 2A.

(B) A close view, in the same orientation as in (A), of the polar network connecting the C-terminal fragment with its two adjacent kinase domains. The Cα atoms of residues comprising the polar network are displayed as spheres. Four selected residues in the network (Lys822, Lys828, Asp988, Asp990; their side chains displayed as sticks) are buried in the core of the kinase/C-terminal fragment interface, suggesting that they play a key role in complex stabilization (Sheinerman et al., 2000). Solid pink lines connect the Cα (or nearest neighbors) atoms of residues that form ion pairs and hydrogen bonds in the network. By symmetry, identical interactions connect residues between the second C-terminal fragment and the kinase domains (not shown). Each residue in the network is involved in a few interactions with neighboring residues. For instance, Asp990, located on the C-terminal domain, interacts with Lys822, located on one kinase domain monomer (yellow), and with Lys799, located on the second kinase domain monomer (light purple), presumably stabilizing the complex.

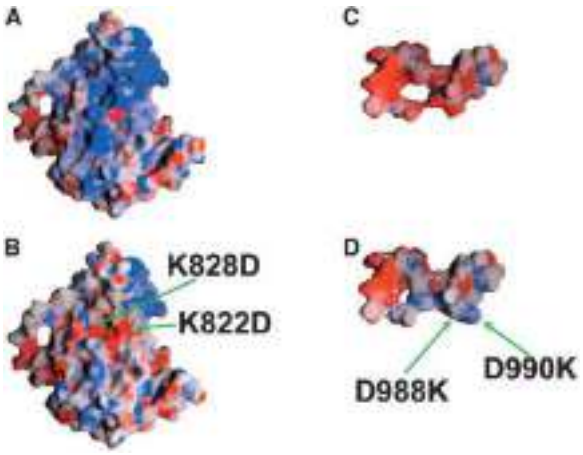


Figure 6. Electrostatic Analysis of Novel Mutations in the EGFR Kinase and C-Terminal Domains

A projection of the electrostatic potential (ϕ) onto the molecular surface of the kinase domain and the C-terminal fragment; the color coding is as in Figures 2B–2D.

(A) The native kinase domain in the same orientation as in Figure 2D.

(B) The R822D/R828D double mutant EGFR kinase domain in the same orientation as in (A).

(C) The native C-terminal fragment of the EGFR rotated 90° to the left relative to its orientation in Figure 2C.

(D) The D988R/D990R double mutant C-terminal domain in the same orientations as in (C).

complex can be tested experimentally by using the double mutant cycle approach (Serrano et al., 1990). Briefly, if the additive effects of mutating two residues separately (e.g., Lys822→Asp and Asp990→Lys) is significantly different from the effect of mutating the same two residues simultaneously, then the two positions are interdependent (Serrano et al., 1990), e.g., are involved in a salt bridge. Based on our analysis of the network, we suggest using a double mutant cycle, in which each step involves mutating a pair of similarly charged residues in the EGFR interface, as specified above.

Model of C-Terminal Domain Regulation of Kinase Activity

One of the phosphorylation sites of the C-terminal domain (Tyr992) is located on the fragment that forms contact with the kinase domain and is therefore inaccessible to phosphorylation in this conformation. The catalytic sites in the kinase domains are facing away from each other in the complex; therefore, transphosphorylation of residues on the kinase domain is improbable. The above two observations suggest that the EGFR crystal structure represents an inactive form of the receptor. The EGFR participates in imperative cell processes and ought to remain inactive under most physiological conditions (Huse and Kuriyan, 2002). Therefore, its inactive state should be very stable. Indeed, the complex in the crystal structure of the EGFR appears to be stable, based on the geometric and charge complementarity, further supporting the notion that this complex is inactive. It has been suggested that ligand-induced contact formation of the extracellular domains would lead to reorientation of the transmembrane domains (Fleishman et al., 2002; Jiang and Hunter, 1999; Moriki et al., 2001) and, subsequently, to rearrangements in the cytoplasmic domains (Figure 1). Any reorganization of the cytoplasmic complex, followed by a change in the position of the negatively charged C-terminal fragment, would lead to electrostatic repulsion between the two positively charged kinase monomers (Figure 2). Hence, this model of conformational changes during receptor activation may constitute a hitherto unknown mode of regulation.

Strong reinforcement of this model of regulation is provided by data on the EGFR analog c-ErbB (Chang et al., 1995). Deletions of a C-terminal fragment of this receptor (corresponding to residues 966–1006 of the EGFR) lead to higher autokinase activity compared to normal c-ErbB and transforming ability in vitro and in vivo. Moreover, a mutant in which the four consecutive acidic residues EEED were replaced by the polar segment QQQN showed higher autokinase activity and a partial transformation phenotype. Since the two mutants and normal receptors have similar rates of degradation, the higher transforming ability of the mutants could not be attributed to a longer half-life of the mutant receptor (Chang et al., 1995). These data are consistent with our results. The four acidic residues, which correspond to the DEED segment (Asp979–Asp982) in the EGFR, are located on the C-terminal fragment (Figure 3) that forms contact with the kinase domain. Our analysis showed that these positions contribute significantly to the negative electrostatic potential of the fragment (Figure 2), and their substitution with polar residues reduces the complementarity between the kinase and C-terminal domains (Figure 4B), presumably destabilizing the inactive complex.

Internal deletions of segments in the C-terminal domain of the EGFR have also been detected in naturally occurring EGFR mutants, which display tumorigenic properties. For example, an internal deletion of residues 959–1030 has been detected in EGFRs sequenced from human glioblastomas (Boerner et al., 2003; Chang et al., 1995; Frederick et al., 2000). Some viral ErbBs contain an in-frame deletion of 139 residues within the intracellular region, immediately following the kinase domain (Boerner et al., 2003; Chang et al., 1995; Frederick et al., 2000). This region contains the C-terminal fragment contacting the kinase domain according to the X-ray structure (Stamos et al., 2002). Our model suggests that the internal deletions in the C-terminal domain yield constitutively active forms of EGFR by means of destabilization of the inactive complex.

Evolutionary Conservation Analysis

The kinase domain of ErbB3 has no catalytic activity, yet it dimerizes with other members of the ErbB family to produce heterodimers with highly efficient catalytic activity (Schlessinger, 2000). These distinct features are manifested in the evolutionary-conservation analysis. ErbB3's kinase domain displays variations in the catalytic site in comparison to other members of the ErbB family, thus rendering it inactive. However, the interface between the kinase domain and the C-terminal fragment

is highly conserved within the ErbBs and their orthologs, including ErbB3. As a reference, an analysis of 121 kinase domains from various TKs showed that the catalytic site, including the ATP and substrate binding loop, was highly conserved, whereas the interface between the kinase domain and the C-terminal fragment was highly variable (data not shown).

Overall, the conservation analysis provides further support for the suggestion that the dimeric complex observed in the crystal structure is not common to all the TKs. However, the contact area between the kinase and C-terminal domains in this complex is common to the ErbBs, which thus maintain the ability to produce homo- and heterodimers through the same interface.

A Network of Correlated Amino Acid Substitutions between Regulatory Elements

By and large, all TKs carry out the same catalytic process. Thus, key residues in the kinase domain, which are responsible for catalysis of phosphotransfer, are under strong evolutionary constraint, as mentioned above. However, in order for the kinases to be involved in numerous and distinct signal transduction pathways, each kinase family exhibits variations in its amino acid sequence that are necessary for the modification of the mode of regulation. Since multiple positions are involved in determining these traits, these sequence variations should occur concomitantly in relevant regulatory elements. In other words, during evolution, substitutions of one residue in regulatory elements may be compensated by a concurrent change in another residue, in order to maintain the structural or functional relationship between these positions (Fleishman et al., 2004b).

In order to look for particular positions that could play a role in regulation, we analyzed the set of 121 multiply aligned TKs of diverse families in search of pairs of amino acid positions that might be evolutionarily correlated (Fleishman et al., 2004b). The analysis revealed 152 pairs of correlated residues, among which we identified a network of 14 highly intercorrelated positions (Figure 7A and Table 1).

The kinase domain includes several regulatory elements, such as the αC helix and activation loop, which play a role in allosteric regulation and are responsible for conformational changes. These elements function together to control activation, i.e., their movements are concurrent and their conformations are mutually dependent (Huse and Kuriyan, 2002). Our analysis showed pairs of evolutionarily correlated positions in these known regulatory elements. For example, Ala743, which is located on the αC helix, is correlated with Gly849 of the activation loop (Figure 7A).

The LVI segment (residues 955–957) of EGFR and its equivalent segments in other ErbBs are necessary for ligand-independent dimerization of the EGFR intracellular domains and for transphosphorylation in ErbB2 and ErbB3 heterodimers through allosteric regulation (Stamos et al., 2002). Leu955 in this LVI segment is correlated with Tyr740, which is located on the αC helix (Figure 7A). The association of the αC helix with a known dimerization motif exemplifies interdomain relationships between regulatory elements in the ErbBs. Both of these

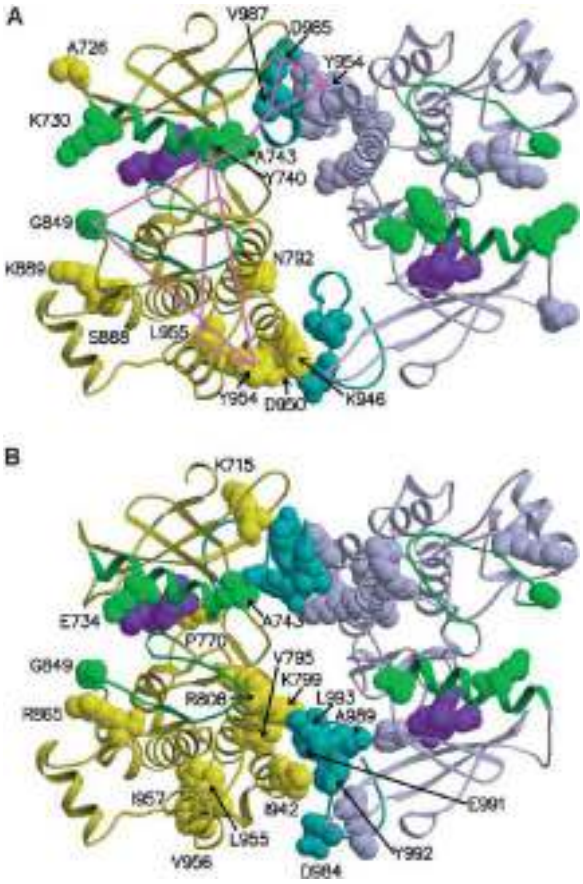


Figure 7. Evolutionarily Correlated and Specificity-Determining Amino Acid Sites

The EGFR homodimeric complex (Stamos et al., 2002) as viewed with an upward rotation of about 180° compared to Figure 2A. The kinase domains, presented by using trace models, are colored light purple and yellow, the C-terminal fragments are colored cyan, and the inhibitor is shown as a purple space-filled model. The αC helix (residues 729–744) and the activation loop (residues 831–852) are colored green.

(A) The residues in the cluster of the most significant pairs of correlated amino acid sites are displayed as space-filled models. Solid pink lines connect a few of the pairs of correlated residues (highlighted in Table 1) in the EGFR homodimer. Correlations within the kinase domain are demonstrated only on the left monomer, and correlations between the kinase and the C-terminal domains are demonstrated only on the upper interface. The correlations between known regulatory elements, such as the αC helix and the activation loop and the interface between the kinase domain and C-terminal fragment, suggest that the latter may also be involved in regulation.

(B) The main specificity-determining residues are located on the αC helix, the activation loop, the C-terminal fragment, and its interfaces on the kinase. This suggests that the regulatory elements in the EGFR had evolved specifically to stabilize the active conformation. Concurrently, an alternative negative regulatory mechanism had evolved in the form of the inactive complex between the kinase and the C-terminal domains. The figures were made by using MOLSCRIPT (Kraulis, 1991) and Raster3D (Merritt and Bacon, 1997).

domains are important for regulation (Huse and Kuriyan, 2002; Stamos et al., 2002); for example, mutations in Leu955 or Tyr740 severely impaired the kinase activity of the EGFR (Stamos et al., 2002; Walker et al., 1998).

Based on these results, we concluded that this net-

Table 1. Correlated Pairs in the TK Family

Pairs of Correlated Positions	Correlation Coefficients
Ala726-Lys730	0.65 (0.49, 0.78)
Ala726-Ser888	0.60 (0.34, 0.77)
Ala726-Lys946	0.45 (0.20, 0.64)
Lys730-Ser888	0.59 (0.37, 0.76)
Tyr740-Tyr954	**0.45 (0.19, 0.67)**
Tyr740-Leu955	**0.48 (0.21, 0.69)**
Ala743-Asn792	0.50 (0.19, 0.76)
Ala743-Gly849	**0.47 (0.18, 0.72)**
Ala743-Ser888	0.45 (0.21, 0.69)
Ala743-Lys889	0.51 (0.14, 0.73)
Ala743-Val987	0.54 (0.34, 0.72)
Asn792-Gly849	0.48 (0.26, 0.70)
Asn792-Ser888	0.50 (0.25, 0.72)
Gly849-Tyr954	**0.56 (0.29, 0.76)**
Ser888-Lys946	0.61 (0.40, 0.78)
Lys946-Asp950	0.52 (0.31, 0.69)
Asp950-Tyr954	0.48 (0.31, 0.65)
Asp950-Leu955	0.50 (0.26, 0.67)
Tyr954-Leu955	**0.54 (0.32, 0.72)**
Tyr954-Asp985	**0.50 (0.30, 0.67)**
Tyr954-Val987	**0.52 (0.36, 0.67)**
Asp985-Val987	**0.52 (0.36, 0.67)**

A list of 22 pairwise correlations between positions comprising the most significant cluster of correlated residues. The trimmed means in the 95% confidence interval of correlations (r), which were calculated from 400 bootstrapping samples, are indicated, and the 95% confidence interval is shown in parentheses (see the Supplemental Data). The numbering of the positions is done according to the EGFR sequence. The pairs of positions that are located on known regulatory regions are highlighted in bold and are connected by solid pink lines in Figure 7A.

work of correlations identified amino acids playing a role in regulation. Interestingly, the same cluster also displays correlations between residues mediating contact between the kinase and C-terminal domains. Tyr954 is located on the kinase domain and contacts the C-terminal fragment. This residue is in close proximity to, and is highly correlated with, residues Asp985 and Val987 of the C-terminal fragment (Figure 7A). Taken together, these correlations consolidate our hypothesis that the contact between the kinase and C-terminal domains is biologically meaningful.

The same cluster of 14 highly intercorrelated positions also includes correlations between positions at the interface of the kinase domain and the C-terminal fragment and known regulatory elements. For instance, Tyr954 located on this interface is correlated with Gly849 of the activation loop, with Leu955 of the LVI segment, and with Tyr740 of the αC helix (Figure 7A). This network of correlations suggests that this interface is also involved in regulation.

Val987 of the C-terminal fragment is correlated with Ala743, which is located on the αC helix (Figure 7A). In this context, it is important to note that the C-terminal domain is a vital modulator of TKs' activity (Jorissen et al., 2003; Schlessinger, 2000), as was elaborated above. For example, structure determination and mutagenesis experiments have shown that the kinase domains of the insulin, the Tie2, and the platelet-derived growth factor β receptor (PDGFR) TKs are autoinhibited by their C-terminal domains through direct contacts with the kinase domain (Chiara et al., 2004; Noelle et al., 2000; Shewchuk et al., 2000). Accordingly, evolutionary correlation between the kinase and C-terminal domains is expected to be general. The mechanism by which direct contacts control activation may vary between the kinases and could not be inferred from the evolutionary-correlation analysis. We anticipate that in the ErbBs, the direct contact between the kinase and C-terminal domains regulates catalysis by the formation of the inactive dimer shown in Figure 2A.

Specificity Determinants in Regulatory Regions

Although TKs share an identical catalytic mechanism, each kinase family is regulated by various means, responds to different ligands, and activates diverse substrates. It is anticipated that certain positions would be responsible for these different traits, and would be reflected in their patterns of substitution (Fleishman et al., 2004a). Due to such differences in functions, these positions are not expected to be strictly conserved in evolution. Rather, they should be conserved among kinases of similar functions in different species (orthologs), and would differ in paralogs. Substitutions involving these residues are presumably responsible for certain alterations in the functions of the various families of the TK superfamily.

We have identified some of these specificity-determining amino acid positions in a set of 121 multiply aligned TKs. The main specificity-determining residues are presented in Figure 7B, and their locations are indicated in Table 2. The list includes residues from the known regulatory regions, as well as residues that connect the kinase and the C-terminal domains and participate in the polar network across the interface (Figure 5B).

Discussion

ErbBs are structurally unique among TKs, as all of the catalytic elements in their kinase domains are ready for phosphotransfer at all times (Stamos et al., 2002). Yet, various functional assays show them not to be constitutively active (Schlessinger, 2000). The absence of a central regulatory module raises a fundamental dilemma, namely, what prevents the receptors from being spuriously activated? One possible mechanism is that changes in the relative orientation of the subunits within a dimer control activation, as suggested by the model of rotation-coupled activation (Jiang and Hunter, 1999). According to this view, contact formation between the extracellular domains leads to reorientation in the transmembrane domain, which is propagated into the cytoplasm (Fleishman et al., 2002; Jiang and Hunter, 1999; Moriki et al., 2001). Thus, the reorientation of the kinase domains vis-à-vis each other serves as a molecular switch that turns the kinase domains "on." What might be the mechanism by which this reorientation is translated into kinase activation is not yet clear.

Understanding the molecular details of how the ErbB proteins are regulated will most probably have to await the emergence of a structure of the full-length receptor in oligomeric complexes. The structures of parts of the kinase and the extracellular domains available today only provide a fragmentary view of the regulatory ele-

Table 2. Specificity Determinants in the TK Family

Position	Correlation Coefficients	Location and Putative Functional Role in the EGFR
Lys715	0.23 (0.11, 0.36)	Located on the kinase at the large interface with the C-terminal fragment; participates in the polar network across the interface (Figure 5).
Pro770	0.24 (0.07, 0.39	Located on the kinase at the large interface with the C-terminal fragment.
Val795	0.29 (0.17, 0.40)	Located on the kinase at the small interface with the C-terminal fragment.
Ile942	0.24 (0.13, 0.36)	Located on the kinase at the small interface with the C-terminal fragment.
Lys799	0.24 (0.10, 0.36)	Located on the kinase at the small interface with the C-terminal fragment; participates in the polar network across the interface (Figure 5).
Glu734	0.27 (0.12, 0.42)	Located on the αC helix of the kinase domain; involved in regulation.
Ala743	0.29 (0.14, 0.42)	Located on the αC helix of the kinase domain; involved in regulation.
Arg808	0.23 (0.12, 0.35)	Located on the kinase domain, close to the activation loop. Involved in hydrogen bonds that stabilize the activation loop (Stamos et al., 2002).
Arg865	0.26 (0.10, 0.41)	Located on the kinase domain, close to the activation loop. Involved in hydrogen bonds that stabilize the activation loop (Stamos et al., 2002).
Gly849	0.28 (0.18, 0.37)	Located on the activation loop of the kinase domain; involved in regulation.
Leu955	0.29 (0.13, 0.42)	A part of the "LVI motif". Important for dimerization of the kinases.
Val956	0.29 (0.17, 0.42)	A part of the "LVI motif". Important for dimerization of the kinases.
Ile957	0.25 (0.08, 0.39)	A part of the "LVI motif". Important for dimerization of the kinases.
His964	0.29 (0.15, 0.43)	A putative negative regulator of EGFR's activity; located on the C-terminal domain.
Leu965	0.24 (0.10, 0.39)	A putative negative regulator of EGFR's activity; located on the C-terminal domain.
Ser967	0.23 (0.10, 0.35)	A putative negative regulator of EGFR's activity; located on the C-terminal domain.
Pro968	0.26 (0.13, 0.40)	A putative negative regulator of EGFR's activity; located on the C-terminal domain.
Ser971	0.27 (0.13, 0.38)	A putative negative regulator of EGFR's activity; located on the C-terminal domain.
Tyr974	0.30 (0.16, 0.43)	A putative negative regulator of EGFR's activity; located on the C-terminal domain.
Asp984	0.31 (0.19, 0.44)	A putative negative regulator of EGFR's activity; located on the C-terminal fragment; participates in the polar network across the interface (Figure 5).
Ala989	0.25 (0.13, 0.38)	A putative negative regulator of EGFR's activity; located on the C-terminal fragment; participates in the polar network across the interface (Figure 5).
Glu991	0.25 (0.11, 0.41)	A putative negative regulator of EGFR's activity; located on the C-terminal fragment; participates in the polar network across the interface (Figure 5).
Leu993	0.29 (0.17, 0.41)	A putative negative regulator of EGFR's activity; located on the C-terminal fragment; participates in the polar network across the interface (Figure 5).
Tyr992	0.24 (0.11, 0.38)	An autophosphorylation site, located on the C-terminal fragment; participates in the polar network across the interface (Figure 5).

A list of 24 out of 47 residues that were identified as specificity determinants (Fleishman et al., 2004a). The location of each residue in the EGFR sequence and its functional role are indicated. The trimmed means in the 95% confidence interval of correlations (r), which were calculated from 400 bootstrapping samples, are indicated, and the 95% confidence interval is shown in parentheses (see the Supplemental Data). In addition to the residues presented above, the list of specificity determinant includes the following residues: V750, Q763, L775, E780, D783, N792, V821, Q825, T830, S888, K889, I899, S901, I902, P910, K925, S933, D950, Q952, Q958, G959, D960, and E961. Their putative roles remain to be tested experimentally.

ments in the structure. In Figure 1, we suggest a model for such regulation in the ErbB family; this model is based on the available structures and is supported by a large body of biochemical and physiological data.

The role of the C-terminal domain as a modulator of kinase activity has been discussed extensively (Cadena et al., 1994; Jorissen et al., 2003), especially in the v-ErbB products (Boerner et al., 2003). Our results offer a model of the molecular mechanism for this modulation (Figure 1). In the inactive state (Figure 1, left), the EGFR extracellular domains assume a tethered structure (Ferguson et al., 2003) that hinders contact formation between the two subunits (Burgess et al., 2003). In this conformation, the extracellular domains are connected to the transmembrane helices in their inactive state (Fleishman et al., 2002), thereby maintaining the intracellular domains as a stable, inactive dimer (Figure 2A). In this state, the C-terminal domain is in contact with the kinase domain and is inaccessible to downstream substrates (Cadena et al., 1994). Ligand-induced activation of the EGFR (Figure 1, right) leads to conformational changes in the extracellular domains, allowing contact formation between the two subunits (Ogiso et al., 2002), followed by a rotation of the transmembrane helices toward their active state (Fleishman et al., 2002; Jiang and Hunter, 1999; Moriki et al., 2001). This switch in the orientation of the transmembrane helices leads to the destabilization of the inactive intracellular dimer. The C-terminal domain detaches from the kinase domain and may undergo phosphorylation, making the kinase accessible to its substrates (Moriki et al., 2001).

The structure of the GW572016 bound EGFR comprises the kinase domain and part of the C-terminal domain that is packed along the kinase domain. In this structure, the C-terminal domain partly blocks the ATP binding site (Wood et al., 2004), as in the inactive forms of the myosin light chain kinase of the Ser/Thr kinase family (Huse and Kuriyan, 2002) and the Tie2 RTK (Shewchuk et al., 2000). That the GW572016 bound EGFR structure shows an inactive conformation that is not primed for catalysis suggests that activation of the EGFR involves conformational changes within the kinase domain, in contrast to the view that the kinase domain is constitutively ready for phosphotransfer (Stamos et al., 2002). We note, however, that the new structure suggests an important role for the C-terminal domain in stabilizing an inactive conformation of the kinase domain (Wood et al., 2004); this finding is in harmony with the model of activation suggested here.

The proposed molecular model may explain the un-

derlying molecular causes of malignancy mediated by EGFRs that contain mutations in their C-terminal domain. According to the model, the transforming properties of these mutations (Boerner et al., 2003; Chang et al., 1995; Frederick et al., 2000) are due to destabilization of the inactive EGFR.

All TKs catalyze the same reaction, which is the transfer of the γ-phosphate of ATP to the hydroxyl group of tyrosine. Indeed, the active conformation of the kinase domain of most TKs is nearly identical. In contrast to the uniform active conformation, TKs differ from each other in their inactive conformations (Huse and Kuriyan, 2002). In some RTKs, as in the PDGFR family, the juxtamembrane domain serves to block the active conformation. Autophosphorylation of tyrosine residues in highly conserved juxtamembrane motifs, specific to each family, relieves autoinhibition (Griffith et al., 2004). In the case of the EGFR family, inhibition by the juxtamembrane domain is less likely, since there are no tyrosine residues in the juxtamembrane segment that can be phosphorylated.

Various regulatory mechanisms could play an important role in ensuring the signaling specificity in the TK superfamily. Accordingly, we suggest that certain amino acid substitutions in regulatory elements were sustained during evolution, leading to alterations in the regulatory mechanisms. This hypothesis is supported by the analysis of specificity determinants (Figure 7B). In the vast majority of the TKs, kinase activity is regulated through a change in the conformation of the activation loop and αC helix. Nevertheless, these regulatory regions undergo different conformational changes in different isoforms, and their inactive conformations are stabilized by fastidious means specific to each kinase family (Huse and Kuriyan, 2002). The ErbBs are further exceptional among TKs, in that the activation loop and αC helix are constitutively stable in the active conformation (Stamos et al., 2002). Our analysis of correlated mutations (Figure 7A) suggests that in order to complement the role of these known regulatory elements in maintaining an active conformation, other residues in ErbBs have evolved to keep the enzyme dormant, as in the "inactive" complex shown in Figure 2A.

We propose that members of the EGFR family utilize the unique regulatory mechanism that is presented in Figure 1. These receptors contain a long C-terminal domain that is involved in signal transmission inside the cell and is also an inherent regulator of kinase activity (Chang et al., 1995). Our results suggest that the complex between the kinase and C-terminal domains of Figure 2A is stable and biologically significant, as indicated by the large intersubunit interface, the electrostatic and geometric complementarity between the C-terminal segments and the kinases (Figures 2B–2D and 5), as well as the evolutionary correlation between specified amino acid sites (Figure 7A). This complex appears to correspond to the basal, inactive form of the receptor, as delineated above and in accordance with previous experimental data (Boerner et al., 2003; Chang et al., 1995). Although our computational analysis and the experimental data support the presence of an inactive dimer (Yu et al., 2002) and the necessity of contact between the kinase and C-terminal domains (Chang et al., 1995), the biological relevance of the crystal dimer has yet to be determined. The importance of the interface between the kinase domain and the C-terminal fragment for the regulation of EGFR activity can be tested experimentally, as delineated in the section entitled "A Network of Ion Pairs and Hydrogen Bonds at the Interface."

Our model of EGFR's regulation (Figure 1) and its relevance to cancer could be further tested by examining the properties of a short peptide analog to the C-terminal fragment. Such a peptide may have a regulatory effect on EGFR activation. For instance, in tumorigenic cells, the short peptide may associate with the kinase domain instead of the truncated C-terminal domain. This would stabilize the inactive configuration and thereby thwart the constitutive activation of the mutant receptor. Interestingly, a similar approach was applied successfully in a recent study on the PDGFR, which is also selfinhibited by direct contact with its C-terminal domain (Chiara et al., 2004). In this work, the authors showed that a soluble peptide, corresponding to the inhibitory fragment in the PDGFR C-terminal domain, delayed the activation of the receptor and inhibited the enhanced kinase activity of a C-terminal truncated PDGFR. Hence, the small peptide mimicked the role of the C-terminal fragment in regulating kinase activity (Chiara et al., 2004). It will be interesting to examine the therapeutic properties of such a peptide in the case of the EGFR.

Experimental Procedures

Biophysical and Structural Analysis

Electrostatic, solvent-accessible surface area calculations and homology modeling were carried out as described in the Supplemental Data.

Collection of Sequence Homologs and Their Alignment

A multiple-sequence alignment (MSA) of homologous kinase domains was produced by combining multiple-structure and sequence alignments to obtain high-quality alignments as described by Al-Lazikani et al. (2001) and in the Supplemental Data. This resulted in an MSA of 121 homologous sequences comprising the kinase domain and about 50 positions C-terminal to it (corresponding to positions 683–998 of the EGFR). The MSA is shown in Supplemental Figure S1 in the Supplemental Data.

Evolutionary Conservation

Evolutionary conservation scores were calculated by using the MSA and *Rate4Site*'s maximum-likelihood algorithm (Pupko et al., 2002), as implemented in the ConSurf web server (Glaser et al., 2003) (http://consurf.tau.ac.il/).

Correlated Amino Acid Substitutions

Pairs of amino acids that appear to change concomitantly during evolution within the TKs were detected by using the MSA and the *CorrMut* algorithm (Fleishman et al., 2004b). The methodological details are provided as Supplemental Data.

Specificity Determinants

Residues in the TK superfamily, which may be responsible for determining specific characteristics in different kinase families, were detected by using the MSA and the *SpecDet* algorithm (Fleishman et al., 2004a). A description of the algorithm is provided as Supplemental Data.

Supplemental Data

Supplemental Data including analysis of the electrostatic potential of representative TKs of known structure; solvent-accessible surface area calculations and homology modeling of selected TKs;

the MSA of the TK family; methodological details of the correlated mutations analysis; and a description of the algorithm used for detecting the specificity-determining residues are available at http://www.structure.org/cgi/content/full/12/12/2265/DC1/.

Acknowledgments

We thank Tony Hunter, Antony Burgess, Joseph Schlessinger, Idit Kopatz, Amit Kessel, and Saul Yankofsky for their critical comments on the manuscript, Miriam Eisenstein for her help in the identification of the dimeric conformation of the EGFR kinase, and Lisa Shewchuk for sharing the coordinates of the EGFR/GW572016 structure before their release. This study was supported by a grant from the Israel Cancer Association (ICA) and by a Research Career Development Award from the Israel Cancer Research Fund (ICRF) to N.B.-T. S.J.F. was supported by a doctoral fellowship from the Clore Israel Foundation. M.L. was supported by a Travel Scholarship from the Constantiner Institute for Molecular Genetics.

Received: June 16, 2004
Revised: September 22, 2004
Accepted: October 8, 2004
Published: December 7, 2004

References

Al-Lazikani, B., Sheinerman, F.B., and Honig, B. (2001). Combining multiple structure and sequence alignments to improve sequence detection and alignment: application to the SH2 domains of Janus kinases. Proc. Natl. Acad. Sci. USA *98*, 14796–14801.

Biswas, R., Basu, M., Sen-Majumdar, A., and Das, M. (1985). Intrapeptide autophosphorylation of the epidermal growth factor receptor: regulation of kinase catalytic function by receptor dimerization. Biochemistry *24*, 3795–3802.

Boerner, J.L., Danielsen, A., and Maihle, N.J. (2003). Ligand-independent oncogenic signaling by the epidermal growth factor receptor: v-ErbB as a paradigm. Exp. Cell Res. *284*, 111–121.

Burgess, A.W., Cho, H.S., Eigenbrot, C., Ferguson, K.M., Garrett, T.P., Leahy, D.J., Lemmon, M.A., Sliwkowski, M.X., Ward, C.W., and Yokoyama, S. (2003). An open-and-shut case? Recent insights into the activation of EGF/ErbB receptors. Mol. Cell *12*, 541–552.

Cadena, D.L., Chan, C.L., and Gill, G.N. (1994). The intracellular tyrosine kinase domain of the epidermal growth factor receptor undergoes a conformational change upon autophosphorylation. J. Biol. Chem. *269*, 260–265.

Chang, C.M., Shu, H.K., Ravi, L., Pelley, R.J., Shu, H., and Kung, H.J. (1995). A minor tyrosine phosphorylation site located within the CAIN domain plays a critical role in regulating tissue-specific transformation by erbB kinase. J. Virol. *69*, 1172–1180.

Chantry, A. (1995). The kinase domain and membrane localization determine intracellular interactions between epidermal growth factor receptors. J. Biol. Chem. *270*, 3068–3073.

Chiara, F., Bishayee, S., Heldin, C.H., and Demoulin, J.B. (2004). Autoinhibition of the platelet-derived growth factor beta-receptor tyrosine kinase by its C-terminal tail. J. Biol. Chem. *279*, 19732–19738.

Cho, H.S., Mason, K., Ramyar, K.X., Stanley, A.M., Gabelli, S.B., Denney, D.W., Jr., and Leahy, D.J. (2003). Structure of the extracellular region of HER2 alone and in complex with the Herceptin Fab. Nature *421*, 756–760.

Dancey, J.E. (2004). Predictive factors for epidermal growth factor receptor inhibitors—The bull's-eye hits the arrow. Cancer Cell *5*, 411–415.

Ferguson, K.M., Berger, M.B., Mendrola, J.M., Cho, H.S., Leahy, D.J., and Lemmon, M.A. (2003). EGF activates its receptor by removing interactions that autoinhibit ectodomain dimerization. Mol. Cell *11*, 507–517.

Fleishman, S.J., Schlessinger, J., and Ben-Tal, N. (2002). A putative molecular-activation switch in the transmembrane domain of erbB2. Proc. Natl. Acad. Sci. USA *99*, 15937–15940.

Fleishman, S.J., Unger, V.M., Yeager, M., and Ben-Tal, N. (2004a). A C-alpha model for the transmembrane alpha-helices of gap-junction intercellular channels. Mol. Cell *15*, 879–888.

Fleishman, S.J., Yifrach, O., and Ben-Tal, N. (2004b). An evolutionarily conserved network of amino acids mediates gating in voltage-dependent potassium channels. J. Mol. Biol. *340*, 307–318.

Frederick, L., Wang, X.-Y., Eley, G., and James, C.D. (2000). Diversity and frequency of epidermal growth factor receptor mutations in human glioblastomas. Cancer Res. *60*, 1383–1387.

Gadella, T.W., Jr., and Jovin, T.M. (1995). Oligomerization of epidermal growth factor receptors on A431 cells studied by time-resolved fluorescence imaging microscopy. A stereochemical model for tyrosine kinase receptor activation. J. Cell Biol. *129*, 1543–1558.

Gamett, D.C., Tracy, S.E., and Robinson, H.L. (1986). Differences in sequences encoding the carboxyl-terminal domain of the epidermal growth factor receptor correlate with differences in the disease potential of viral erbB genes. Proc. Natl. Acad. Sci. USA *83*, 6053–6057.

Glaser, F., Pupko, T., Paz, I., Bell, R.E., Bechor-Shental, D., Martz, E., and Ben-Tal, N. (2003). ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. Bioinformatics *19*, 163–164.

Gotoh, N., Tojo, A., Hino, M., Yazaki, Y., and Shibuya, M. (1992). A highly conserved tyrosine residue at codon 845 within the kinase domain is not required for the transforming activity of human epidermal growth factor receptor. Biochem. Biophys. Res. Commun. *186*, 768–774.

Griffith, J., Black, J., Faerman, C., Swenson, L., Wynn, M., Lu, F., Lippke, J., and Saxena, K. (2004). The structural basis for autoinhibition of FLT3 by the juxtamembrane domain. Mol. Cell *13*, 169–178.

Huse, M., and Kuriyan, J. (2002). The conformational plasticity of protein kinases. Cell *109*, 275–282.

Jiang, G., and Hunter, T. (1999). Receptor signaling: when dimerization is not enough. Curr. Biol. *9*, R568–R571.

Jones, S., and Thornton, J.M. (1996). Principles of protein-protein interactions. Proc. Natl. Acad. Sci. USA *93*, 13–20.

Jorissen, R.N., Walker, F., Pouliot, N., Garrett, T.P., Ward, C.W., and Burgess, A.W. (2003). Epidermal growth factor receptor: mechanisms of activation and signalling. Exp. Cell Res. *284*, 31–53.

Kraulis, P.J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J. Appl. Crystallogr. *24*, 946–950.

Massoglia, S., Gray, A., Dull, T.J., Munemitsu, S., Kun, H.J., Schlessinger, J., and Ullrich, A. (1990). Epidermal growth factor receptor cytoplasmic domain mutations trigger ligand-independent transformation. Mol. Cell. Biol. *10*, 3048–3055.

Merritt, E.A., and Bacon, D.J. (1997). Raster3D photorealistic molecular graphics. Methods Enzymol. *277*, 505–524.

Moriki, T., Maruyama, H., and Maruyama, I.N. (2001). Activation of preformed EGF receptor dimers by ligand-induced rotation of the transmembrane domain. J. Mol. Biol. *311*, 1011–1026.

Nicholls, A., Sharp, K.A., and Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins *11*, 281–296.

Noelle, V., Tennagels, N., and Klein, H.W. (2000). A single substitution of the insulin receptor kinase inhibits serine autophosphorylation in vitro: evidence for an interaction between the C-terminus and the activation loop. Biochemistry *39*, 7170–7177.

Ogiso, H., Ishitani, R., Nureki, O., Fukai, S., Yamanaka, M., Kim, J.H., Saito, K., Sakamoto, A., Inoue, M., Shirouzu, M., and Yokoyama, S. (2002). Crystal structure of the complex of human epidermal growth factor and receptor extracellular domains. Cell *110*, 775–787.

Pelley, R.J., Maihle, N.J., Boerkoel, C., Shu, H.K., Carter, T.H., Moscovici, C., and Kung, H.J. (1989). Disease tropism of c-erbB: effects of carboxyl-terminal tyrosine and internal mutations on tissue-specific transformation. Proc. Natl. Acad. Sci. USA *86*, 7164–7168.

Pupko, T., Bell, R.E., Mayrose, I., Glaser, F., and Ben-Tal, N. (2002). Rate4Site: an algorithmic tool for the identification of functional

regions in proteins by surface mapping of evolutionary determinants within their homologues. Bioinformatics *18*, S71–S77.

Raines, M.A., Maihle, N.J., Moscovici, C., Moscovici, M.G., and Kung, H.J. (1988). Molecular characterization of three erbB transducing viruses generated during avian leukosis virus-induced erythroleukemia: extensive internal deletion near the kinase domain activates the fibrosarcoma- and hemangioma-inducing potentials of erbB. J. Virol. *62*, 2444–2452.

Riedel, H., Schlessinger, J., and Ullrich, A. (1987). A chimeric, ligand-binding v-erbB/EGF receptor retains transforming potential. Science *236*, 197–200.

Robinson, H.L., Tracy, S.E., Nair, N., Taglienti-Sian, C., and Gamett, D.C. (1992). Characterization of an angiosarcoma-inducing mutation in the erbB oncogene. Oncogene *7*, 2025–2030.

Schlessinger, J. (2000). Cell signaling by receptor tyrosine kinases. Cell *103*, 211–225.

Schlessinger, J. (2003). SIGNAL TRANSDUCTION: autoinhibition control. Science *300*, 750–752.

Serrano, L., Horovitz, A., Avron, B., Bycroft, M., and Fersht, A.R. (1990). Estimating the contribution of engineered surface electrostatic interactions to protein stability by using double-mutant cycles. Biochemistry *29*, 9343–9352.

Sheinerman, F.B., Norel, R., and Honig, B. (2000). Electrostatic aspects of protein-protein interactions. Curr. Opin. Struct. Biol. *10*, 153–159.

Shewchuk, L.M., Hassell, A.M., Ellis, B., Holmes, W.D., Davis, R., Horne, E.L., Kadwell, S.H., McKee, D.D., and Moore, J.T. (2000). Structure of the Tie2 RTK domain: self-inhibition by the nucleotide binding loop, activation loop, and C-terminal tail. Structure *8*, 1105–1113.

Stamos, J., Sliwkowski, M.X., and Eigenbrot, C. (2002). Structure of the epidermal growth factor receptor kinase domain alone and in complex with a 4-anilinoquinazoline inhibitor. J. Biol. Chem. *277*, 46265–46272.

Walker, F., Kato, A., Gonez, L.J., Hibbs, M.L., Pouliot, N., Levitzki, A., and Burgess, A.W. (1998). Activation of the Ras/mitogen-activated protein kinase pathway by kinase-defective epidermal growth factor receptors results in cell survival but not proliferation. Mol. Cell. Biol. *18*, 7192–7204.

Wedegaertner, P.B., and Gill, G.N. (1992). Effect of carboxyl terminal truncation on the tyrosine kinase activity of the epidermal growth factor receptor. Arch. Biochem. Biophys. *292*, 273–280.

Wood, E.R., Truesdale, A.T., McDonald, O.B., Yuan, D., Hassell, A., Dickerson, S.H., Ellis, B., Pennisi, C., Horne, E., Lackey, K., et al. (2004). A unique structure for epidermal growth factor receptor bound to GW572016 (Lapatinib): relationships among protein conformation, inhibitor off-rate, and receptor activity in tumor cells. Cancer Res. *64*, 6652–6659.

Yarden, Y., and Schlessinger, J. (1987). Epidermal growth factor induces rapid, reversible aggregation of the purified epidermal growth factor receptor. Biochemistry *26*, 1443–1451.

Yarden, Y., and Sliwkowski, M.X. (2001). Untangling the ErbB signalling network. Nat. Rev. Mol. Cell Biol. *2*, 127–137.

Yu, X., Sharma, K.D., Takahashi, T., Iwamoto, R., and Mekada, E. (2002). Ligand-independent dimer formation of epidermal growth factor receptor (EGFR) is a step separable from ligand-induced EGFR signaling. Mol. Biol. Cell *13*, 2547–2557.

APPENDIX III

Experimental Procedures (for Chapter 2) carried out by Katia Herz, Under the Supervision of Etana Padan (The Hebrew University).

Bacterial Strains and Culture Conditions

EP432 is an *Escherichia coli* K-12 derivative, which is *mel*BLid, Δ*nha*A1::*kan*, Δ*nha*B1::*cat*, Δ*lac*ZY, thr1 [1]. Cells were grown either in L broth (LB) or in modified L broth (LBK [2]).

Plasmids

Plasmid pAXH (previously called pYG10), a pET20b (Novagen) derivative, encodes His-tagged NhaA [3]. pCL-AXH, a derivative of pAXH, encodes a His-tagged CL-NhaA [3]. pCL-AXH2, a derivative of pCL-AXH, lacks a BglII site at position 3382 [4]. pCL-AXH3, a derivative of pCL-AXH2, contains a BstXI silent site at position 248 in *nha*A.

Site-Directed Mutagenesis

Site-directed mutagenesis was conducted according to a PCR-based protocol [5] or DpnI-mediated site-directed mutagenesis [6]. For Cys-replacement of W62C, F71C, F72C and G76C and N64C, we used pCL-AXH3 as a template. Mutations H225R and H225C have been previously described [7].

Isolation of Membrane Vesicles and Assay of Na^+/H^+ Antiporter Activity

EP432 cells transformed with the relevant plasmids were grown, and everted vesicles were prepared and used to determine the Na^+/H^+ or Li^+/H^+ antiporter activity as described [8, 9]. Assay of antiporter activity was based on measurement of the Na^+- or Li^+-induced changes in ΔpH, as measured by acridine orange, a fluorescent probe of ΔpH. The fluorescence assay was performed with 2.5 ml of reaction mixture

containing 50−100 μg of membrane protein, 0.5 μM acridine orange, 150 mM KCl, 50 mM BTP, and 5 mM $MgCl_2$, and the pH was titrated with HCl. After energization with D-lactate (2 mM), quenching of the fluorescence was allowed to achieve a steady state, and then Na^+ or Li^+ was added. Reversal of the fluorescence level (dequenching) indicates that protons are exiting the vesicles in antiport with Na^+ or Li^+. As previously shown, the end level of dequenching provides a good estimate of the antiporter activity [10] and the ion concentration that yields half-maximal dequenching gives a good estimate of the apparent K_M of the antiporter [10, 11]. The apparent K_M for Li^+ of EcNhaA is 10-fold lower than that for Na^+ (0.02 mM and 0.2 mM at pH 8.5, respectively). The Na^+/H^+ or Li^+/H^+ antiporter activity was measured with 5 times the concentration of the apparent K_M.

References

1. Pinner, E., Y. Kotler, E. Padan and S. Schuldiner, Physiological role of *nhaB*, a specific Na^+/H^+ antiporter in *Escherichia coli*. *J. Biol. Chem.* (1993) **268**: 1729-1734.
2. Padan, E., N. Maisler, D. Taglicht, R. Karpel and S. Schuldiner, Deletion of *ant* in *Escherichia coli* reveals its function in adaptation to high salinity and an alternative Na^+/H^+ antiporter system(s). *J. Biol. Chem.* (1989) **264**: 20297-20302.
3. Olami, Y., A. Rimon, Y. Gerchman, A. Rothman and E. Padan, Histidine 225, a residue of the NhaA-Na^+/H^+ antiporter of *Escherichia coli* is exposed and faces the cell exterior. *J. Biol. Chem.* (1997) **272**: 1761-1768.
4. Galili, L., A. Rothman, L. Kozachkov, A. Rimon and E. Padan, Trans membrane domain IV is involved in ion transport activity and pH regulation of the NhaA-Na(+)/H(+) antiporter of Escherichia coli. *Biochemistry* (2002) **41**: 609-17.
5. Ho SN, H.H., Horton RM, Pullen JK, Pease LR, Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* (1989) **77**: 51-9.

6. Fisher, C.L., and Pei, G. K., Modification of a PCR-based site-directed mutagenesis method. *Biotechniques* (1997) **23**: 570-574.
7. Gerchman, Y., Y. Olami, A. Rimon, D. Taglicht, S. Schuldiner and E. Padan, Histidine-226 is part of the pH sensor of NhaA, a Na+/H+ antiporter in Escherichia coli. *Proc. Natl. Acad. Sci. USA* (1993) **90**: 1212-6.
8. Rosen, B.P., Ion extrusion systems in *E. coli. Methods Enzymol.* (1986) **125**: 328-386.
9. Goldberg, E.B., T. Arbel, J. Chen, R. Karpel, G.A. Mackie, S. Schuldiner and E. Padan, Characterization of a Na^+/H^+ antiporter gene of *Escherichia coli. Proc. Natl. Acad. Sci. USA* (1987) **84**: 2615-2619.
10. Schuldiner, S. and H. Fishkes, Sodium-proton antiport in isolated membrane vesicles of Escherichia coli. *Biochemistry* (1978) **17**: 706-11.
11. Tsuboi, Y., H. Inoue, N. Nakamura and H. Kanazawa, Identification of membrane domains of the Na^+/H^+ antiporter (NhaA) protein from Helicobacter pylori required for ion transport and pH sensing. *J. Biol. Chem.* (2003) **278**: 21467-21473.

תקציר

המטרה העיקרית של עבודת הדוקטורט שלי הייתה להעשיר את הידע על מנגנוני בקרה של חלבונים המשחקים תפקידי מפתח בתהליכים פיזיולוגיים ובמחלות. להשגת המטרה הזו התמקדתי במבני חלבונים, ושילבתי מידע ממחקרים ביוכימיים וקליניים עם אנליזות חישוביות שביצעתי על חלבונים אילו בהתבסס על ההיסטוריה האבולוציונית שלהם ותכונותיהם הפיזיכוכימיקליות. באופן ספציפי, שאפתי להבנה מנגנונית של הפונקציה והבקרה של שתי משפחות חלבונים חוצי-ממבראנה, משפחת ה- ErbB ומשחלפי Na^+/H^+. שתי משפחות אלה מעורבות במחלות קשות, והעשרת המידע לגבי פעילותם היא בעלת חשיבות קלינית גבוהה.

בתחילה התעמקתי בלנסות לשפוך אור על מנגנון הפעולה של משפחת ה- ErbB. רצפטורי טירוזין קינאזות ממשפחת ה- ErbB משחקים תפקידי מפתח בתהליכים תאיים חיוניים ובמגוון מחלות סרטן. רצפטורים אילו שונים מקינאזות אחרות בכך שהפעלתם אינה תלויה במצב הזירחון שלהם. יותר מכך, המבנה הגבישי הראשון של מתחם הקינאז של החבר הראשון במשפחה, שנקרא גם הרצפטור לפקטור גדילה אפידרמלי (EGFR), הראה שכל האלמנטים מוכנים לקטליזה במצב הלא-מזורחן. לכן, המידע הקיים העלה פרדוקס, היות והיה ידוע שחלבונים אילו אינם פעילים תמידית. בעקבות כך חיפשתי מנגנון בקרה חלופי. היפותזת העבודה שלי הייתה שחלבונים אילו מבוקרים על-ידי מנגנון פנימי למתחם התוך תאי שאינו תלוי בזרחון. התוצאות שלי הראו שהמתחם ה- C טרמינלי של ה- EGFR משמש לדאון-רגולציה של הפעילות הקטליטית דרך מגע ישיר עם מתחם הקינאז. מידע ניסיוני שהתפרסם לאחרונה היווה תמיכה לממצאים אילו.

מאז פרסום הממצאים שלי, כמות גדולה של מידע הצטברה, כולל ממצאים מבניים וביוכימיים. בהתבסס על מידע זה, סיכמתי לאחרונה מודל שכולל הרבה רבדים של בקרה על פעילות ה- EGFR הנאכפים על-ידי המתחמים המבניים השונים שלו, והבקרה ההדדית ביניהם. המודל מספק הסבר, ברמה המולקולארית, לאפקט של מוטציות הגורמות לסרטן, ומציע גישה תרפויטית חדשה למקרי סרטן הקשורים למשפחת ה- ErbB.

בדומה למשפחת ה- ErbB, גם ומשחלפי Na^+/H^+ מעלים שאלות מעניינות בקשר ליחסי מבנה-פעולה בחלבונים חוצי-ממבראנה. חלבונים אליו מובילים יוני נתרן ופרוטונים דרך הממבראנה ומשחקים תפקיד מפתח בשמירת ההומיאוסטסיס בתא, וגם בתהליכים פתולוגים כמו מחלות סרטן ולב. למרות מידע רב ממחקרים הנעזרים במוטציות, המידע המולקולארי על מנגנון שיחלוף הקטיונים עדיין אינו ידוע, בעיקר בגלל שלא קיימים מבנים של טרנספורטרים אילו. על כל פנים, בגלל קשיים טכניים שונים בביטוי וניקוי של חלבונים אילו, נראה שאנחנו עדיין בפיגור של שנים מקביעת מבנה.

המבנה של משחלף Na^+/H^+ מ- *Escherichia coli* (EcNhaA) נקבע לאחרונה בעזרת קריסטלוגרפיית קרני-X. המבנה הייחודי פתח את הדלת להבנת יחסי מבנה-פעולה בטרנספורטרים הומאניים. EcNhaA ומשחלפי Na^+/H^+ אאוקריוטים כנראה חולקים אותה פונקציה וקיפול מבני, ולכן השתמשתי במבנה הגבישי לצפות את המבנה של האיזופורם הראשון של משחלף Na^+/H^+ ההומאני (NHE1). בנית המבנה הייתה מאוד מאתגרת היות וההתאמה הרצפית בין שני החלבונים היא בערך 10%, הרבה מתחת לסף המקובל למידול אמין. לכן שילבתי בין

תוצאות של שיטות של fold-recognition והתאמת רצפים, עם התערבות ידנית. המודל המבני שלי קיבל תוקף בעזרת קריטריונים שונים, כמו פרופיל שימור אבולוציוני ומידע מוטגנטי.

המודל המבני הוא שימושי בעיקר להנחיית ניסיונות ביוכימיים. לפיכך, ביססתי שיתוף פעולה עם פרופסור אתנה פדן מהאוניברסיטה העברית, מי שהייתה מעורבת בקביעת המבנה של ה- EcNhaA. באופן ספציפי, במודל שלי, חומצות-אמינו שמשתתפות בקשירה של מעכבים חשובים-קלינית של NHE1 נמצאות בקרבה אחת לשנייה על המבנה, בעוד שהן מפוזרות על הרצף הראשוני. זה היווה תמיכה חשובה למודל וגם אפשר לי להסיק לגבי המיקום של קשירת המעכב על גבי המבנה של ה- EcNhaA. ניבוי זה קיבל חיזוק מניסיונות מוטגנזה-מכוונת שנעשו על-ידי קטיה הרץ (תחת הנחייתה של אתנה פדן). תוצאת זו מחזקת מאוד את המוטיווציה שלי לצפות את המבנה של החלבון ההומאני NHE1 על בסיס המבנה הגבישי של החלבון החיידקי.

בנוסף, מצאתי ששני הטרנספורטרים חולקים מנגנון שחלוף-קטיונים דומה. זה אפשר לי להסיק אינפורמציה מהתובנות המנגנוניות הקימות עבור החלבון החיידקי על מנת להבין יותר את מנגנון הפעילות של NHE1. לפיכך, שילבתי בין מידע אמפירי והמודל המבני, על מנת להציע מנגנון נגישות-אלטרנטיבית (alternating-access mechanism), ברמה המולקולארית, של שחלוף Na^+/H^+.

העבודה בוצעה בהנחייתו של

פרופסור ניר בן-טל

מחקר חישובי על משפחות חלבונים נבחרות

תזה לקראת תואר "דוקטור לפילוסופיה"

מאת: מיטל לנדאו

הוגש לסנאט של אוניברסיטת תל-אביב

דצמבר, 2007